

23 September 1982

Nothing is even more expensive

The international Atomic Energy Agency (celebrating its 25th anniversary) has a tenable niche on nuclear power but it is not the only custodian of peace.

The outlook for the international nuclear power business seems to have improved a little at last week's meeting of the board of governors of the International Atomic Energy Agency (IAEA) at Vienna. That, at least, is what the public noises mean. For one thing, the authority has been able to spell out in its annual budget a prospectus for a much more meticulous procedure for accounting for fissile material ostensibly under international safeguards that might be used for making bombs. For another, in spite of the deteriorating economics of nuclear power enterprises in countries such as the United States, the economic outlook in countries such as West Germany is not nearly as parlous even with a change of government in the offing. And then the authority had the wit to put up Dr Sigvard Eklund, its director-general until last year, to make the keynote speech at its annual public meeting this week on experience with nuclear power and to echo the late Homi Bhabha in saying "No energy is more expensive than no energy". That message will be remembered.

At the same time, the recognition that countries differ enormously among themselves in their assessment of the importance of nuclear power will begin to sink in. With hindsight, it is entirely credible that, in the United States, the nuclear power industry should have been made uneconomic by the public protests (on the grounds of amenity and safety) of people who are broadly speaking assured of electricity supplies at reasonable cost from conventional fuel sources. In fuel-importing countries such as left-wing France, it is understandable that the balance should lie the other way — and that it should be exaggerated by M. Chevenement's enthusiasm for the future. It is more of a surprise that Mr Helmut Schmidt's chancellorship should have been brought to an end partly on the issue of nuclear power, when West Germany is more dependent on imported fuel than most other European countries except Denmark and Sweden.

The truth, of course, is that the nuclear issue in West Germany springs not from the federal republic's need of energy of some kind (on the grounds that no energy would cost more) but from Mr Schmidt's advocacy of President Carter's fleeting belief that there should be neutron bombs in West Germany. Paradoxically, Mr Schmidt's opponents on the nuclear issue, the wayward "Greens", may be more willing to toe the line when the Christian Democrats are in office. The conventional but cynical wisdom that it takes a left-wing government to accomplish right-wing goals (and the other way round) is undermined by the way in which the expectations of those who oppose the policies of their own governments diminish when their governments are replaced. In the end, what will matter is that the countries most in need of nuclear power (and able to afford it) will build nuclear power stations (and, given their need, will be able to do so cheaply); the others will wait.

The problem for the International Atomic Energy Agency that should occupy its new director, Mr Hans Blix, is therefore what has occupied him in his short tenure of his post — how to arrange that countries buying fissile material from others do not use it for making nuclear weapons or, by threatening to do so, disturb the peace. Outwardly, he must be seen to worry. But the nations most often canvassed as the leading members of the next nuclear generation, Israel for example, live so dangerously already that they are more likely to be constrained by conventional diplomacy than by anything a United Nations agency could accomplish

(witness this week's withdrawal from West Beirut). The trick, for the next year or so, must be to repeat the demonstration in the past year that international safeguards function effectively while showing that they are consistent with, for example, the reprocessing of spent fuel. Mr Blix appears to have the spirit even for that enterprise.

Health service sickness

The British government is heading for a row on nationalized medicine; it should reorganize.

The British National Health Service, variously regarded as a triumph of social engineering and as an embodiment of "socialized medicine", is in trouble. The underlying problems are financial. Since the creation of the health service in 1948, the assumption that the cost of medical care should be a charge on public funds has increasingly been challenged by the disappointing growth of the British economy and by the rapid increase (in real terms) of the cost of medical care, itself a measure of the advance of medical technology. The cost of the National Health Service, £14,500 million in 1982–83, is roughly six per cent of the Gross Domestic Product and has increased by some 5 per cent (in real terms) since 1979–80. While the total cost is less than the proportion of national income spent on health care in, for example, West Germany and the United States, successive British governments of all colours have been increasingly perplexed to know how to pay the bills wished on them three decades ago. Although they have differed in their generosity towards the National Health Service, they seem all to have agreed that the salaries paid to those working in it should be kept as low as possible. The consequences have now caught up with Mrs Thatcher's government.

After decades of parsimony, many health service workers are now paid much less than they would be paid in comparable jobs elsewhere. That trainee nurses earn less than they would be paid if they were unemployed is a relic of the days when nursing was considered a kind of extension of voluntary social work, but the nurses are no longer the genteel profession they once were, and have formally rejected, as would have a trades union, this year's offer of a 7.5 per cent salary increase. Those nurses belonging to the Royal College of Nurses (perhaps 60 per cent of all student and qualified nurses) were prevented by their charter from joining this week's one-day strike by other health service workers. The possibility that the self-denying ordinance may be removed from their charter is, however, a sign that gentility has gone for good. (The decision of the Trades Union Congress to urge that other trades unionists should strike in sympathy for at least an hour is, by contrast, more probably a sign of its eagerness to test the effectiveness of the legislation that makes secondary strikes illegal, and those engaging in them susceptible to civil actions by their employers.) So British governments must steel themselves for annual confrontations over health service pay, for there is no prospect of paying proper salaries to those most deprived while keeping the total cost of the National Health Service commensurate with the likely growth of the British economy. But such a state of affairs could not continue for long without damaging the capacity of the health services to care for the sick;



for those who work for the National Health Service for too little pay, but whose services are essential, will in due course choose to work elsewhere.

The dilemma is familiar, not only in the United Kingdom. Whether medical services are provided free by the state or by some combination of free market and medical insurance, the experience of the past several decades shows that total costs increase more quickly than most other economic indicators. This is only to be expected. The ageing populations of most countries inevitably make greater demands on medical services — but the ageing has itself been made possible by the diligent application of novel medical technology. The old dream of Mr Aneurin Bevan, the architect of the National Health Service, that a comprehensive (and free) health service would reduce the real cost of health care is plainly an illusion. Most probably, the real resources spent by prosperous societies will continue to grow in comparison with other kinds of expenditure (as on food, for example). It is not surprising that the present British government should already be casting round for other ways of financing health services than from direct taxation — and, politics being what they are, that this should have provoked fierce protests from its political opponents. How then should the government proceed?

The time has long since gone when the principle of the National Health Service — free medical care for all — was an ideological issue in the United Kingdom. The past three decades have shown that the mere existence of the National Health Service has been a civilizing and enlightening social influence in Britain. That indigent people and the chronically sick have been cared for well and sympathetically has had a powerfully beneficent influence on the temper of society. The British government, which has been flirting with the notion that medical costs should be to some extent met from insurance schemes, cannot now abandon the principle that medical treatment should be provided for children, the indigent, the chronically sick and the elderly and that, at least for the first three groups, that it should be free, with no questions asked. Once this is accepted, the choice between methods of financing — say from general revenue or health insurance — is logically unimportant. Thus if, as in West Germany, health insurance were compulsory, people with jobs would be unaffected — their insurance premiums would presumably be met out of the income tax they would save. The reality, of course, would be more complicated. Changing the system by which British health services are financed would replace one set of inequities by another, perhaps more burdensome. But changing the system would not of itself improve the quality of medical care.

The now bitter argument about nurses' pay is powerful evidence of that. The immediate complaint is that 7.5 per cent extra is all too modest. Many would agree that qualified nurses should be paid twice as much. The trouble is that such largesse is ruled out not merely by the straitened state of the public finances but also because the nurses' pay is linked with that of the substantial army of ancillary health service workers whose jobs might often just as well be done by private contractors or even by computers. The strongest argument for changing the financial basis of the service as a whole is that there would then be a natural opportunity for looking again at the necessity for this huge workforce — and probably for dispensing with a large part of it.

But why should not the managers of the National Health Service undertake that task off their own bat, without waiting for the financial basis of their operations to be changed? Because there are no managers. Most British nationalized industries are run by public corporations whose members can (in theory) be fired for incompetence. The National Health Service, by contrast, is technically the ultimate responsibility of the Secretary of State for Health and Social Security, who no doubt delegates day to day responsibility to his civil servants. The result is that even administrative issues are, by definition, political questions. If physicians seek a pay increase, or wish to change their terms of service, they must negotiate with identifiable parts of the Whitehall machine, staffed by different officials every few years. Titularly, the hospital service is the responsibility of regional health authorities, large committees consisting of representatives

of the great and the good (retired vice-chancellors are conspicuous) and of regional political interests with chairmen chosen for their neutrality rather than for their managerial skills. The consequence is that radical questions (such as whether the health services would be more efficient if hospitals were financially autonomous, with incomes somehow linked with the numbers of patients treated) can never be considered. In such circumstances, the wonder is not that the British health services are so inefficient (and so badly paid) but that they have survived at all. The present government is right to look for different ways of paying for the same job, but it must also find better ways of arranging that the whole system is properly administered — and not by itself.

Academies not international

How can the International Council of Scientific Unions make best use of the opportunities ahead?

The International Council of Scientific Unions is one of those organizations that would have to be invented if it did not exist but which, once invented, is almost incapable of doing the job its inventors planned. The logic goes like this. National scientific academies are in a particular sense representative of what is best in national scientific communities; they are not elected in a democratic sense, but they perpetuate themselves by electing their own members (who are usually their successors). When, and, if so, because the criteria for election involve intellectual attainment, the resulting national academies can be unexpectedly influential. Over the years, such undemocratic academies have demonstrated that their influence can extend beyond the interests of those who happen to be their members to the interests of the national scientific community at large. So is it not reasonable to expect that a consortium of such academies might exert its collective influence in the interests of the international scientific community, however defined? That is the argument from which ICSU springs.

The calculation is, of course, deficient. And the products of ICSU's work fall far short of what they would be if the calculation were correct, as last week's general assembly at Cambridge showed. That there should have been an argument of any kind about the terms on which the People's Republic of China should belong, and whether Taiwan should have to withdraw if membership were granted on politically unpalatable terms, is absurdly inconsistent with ICSU's ultimate objectives. Why not both? That (see page 292) is how it was decided. The vague form of words on which the compromise is based will not, however, serve as a model for deciding what to do when similar issues arise in more politically contentious international settings. (Chinese membership of most United Nations organizations now habitually takes precedence over that of Taiwan, but the question whether Israel should belong to the European or Middle Eastern section of the World Health Organization was a running sore even before the events of the past few weeks.) International organizations with more clout will debate these forms of words even more exhaustively than at Cambridge last week. ICSU's success on the Chinese question is a measure of its impotence.

ICSU should be worrying more than it does about that state of affairs. Uncontentious good sense (such as the proposal that short-lived radioactive isotopes should be free from import duty) will sweep all before them in due course, but governments (to judge only by the innocent Australian government's performance) are still some way from recognizing that ICSU should have a right to say who from overseas should attend the conferences its member unions sponsor. But is there a chance that ICSU might have more clout if it set about transforming its way of working so as to replicate the ways in which national academies function? For the past several decades, such questions have been overlain by ICSU's liking for paperwork (no doubt acquired by infection from UNESCO, its UN sponsor) and by its lack of funds. Has the time come to break that mould? And does the new leadership have the courage or even the inclination?

Pentagon blocks open exchange

Weinberger in optics meeting censorship

Washington

In the latest crackdown by the Reagan Administration on what it says is a dangerous flow of US scientific information to the Soviet Union, Secretary of Defense Caspar Weinberger last month authorized an eleventh-hour move that prevented approximately 100 scientific papers from being given at a San Diego meeting and caused confusion and even "panic", according to one account, when the meeting was held.

The 26th annual international technical symposium of the Society of Photo-Optical Instrumentation Engineers (SPIE) opened on 18 August on schedule, in spite of the fact that immediately before that, military officials had warned many participants that their papers were too sensitive to be given. The Pentagon had expected 28 Soviet scientists to attend, and it concluded, after checking with the armed services and on the basis of a memorandum to Mr Weinberger, that they represented "a major security problem or disaster". The authors were warned that many papers might not have had proper clearance from the Department of Defense, and about one-sixth of those who had planned to give papers voluntarily withdrew them.

The resulting confusion about clearance procedures led to worried engineers at the meeting badgering military officials to find out whether they would be in trouble if they delivered their papers. Most participants who had carried out their work on defence grants, contracts or subcontracts had cleared their papers through local contracting offices or military bases, in accordance with past procedure. They were apparently unaware of a new rule promulgated in April requiring papers that contain "technical data" of a sensitive nature to be cleared by the Pentagon.

The Pentagon learned of the contents of the meeting only by chance, when an official read a copy of the programme and showed it to Dr Stephen D. Bryen, Deputy Assistant Secretary of Defense for international economics, trade and security policy. "Even if you were a zoologist you would have found that program disturbing", says one military official. "They were going to describe the reconnaissance capabilities of the F-18 fighter and how to see a US guy through the army's obscurants." (Obscurants are smoke or gases used to hide troop movements.) "We didn't withdraw any papers", the official said. "We just warned people they might be in violation of the April rules."

According to the Department of Defense, the authority for the withholding of technical information — regardless of whether or not it should be classified — is the 1954 Arms Export Control Act, as a result of which the State Department issued the International Traffic in Arms Regulations (ITAR) which prohibit the export of certain kinds of equipment or of "technical data" about its use. In addition, the 1979 Export Administration Act calls for development of a list of strategically important technologies and goods that should not go to the Soviet Union.

SPIE is the principal professional group in a field with many military applications. SPIE officials were upset by what

happened at its meeting and by the resignation over the incident of some of its members. Richard Wollensack, a vice-president of the Itek Corporation and president of SPIE, spent three hours talking to defence officials last week. The society stressed its interest in avoiding future problems with the Department of Defense as well as in the continued freedom of its members to present papers at meetings. Military officials stressed that no blame attached to the society.

But the cure for the problem may be more drastic than the disease. The Department of Defense may form a committee of Pentagon officials to monitor more systematically material to be presented at

West German science after Schmidt

Heidelberg

What are the prospects for science and technology in West Germany if a Christian Democratic administration emerges in Bonn? As described in the party's document on research, *Technologischepolitisches Konzept*, the strategy will be to restore the prestige and credibility of research and development in science and technology as keys to future prosperity.

Christian Lenzer, spokesman for the Christian Democratic Union/Christian Social Union (CDU/CSU) on research and technology, emphasizes the need for more basic research, but sponsorship and direction would continue within the existing self-administering framework of the Deutsches Forschungsgemeinschaft, the Max Planck Society, the foundations and large institutions.

In the areas of applied research and development, however, CDU/CSU would change the emphasis from direct to indirect support, thus returning to the pattern of the 1960s for which employers' organizations have been asking.

Industrial research would be primarily self-financed and government support would take the form of indirect financial incentives such as tax concessions (to be assessed by the tax authorities), quicker depreciation of research and development investments, provision of risk capital and tax allowances for patenting and licensing. CDU stresses the importance of free market forces, decentralization and reduction in state control in creating the right climate to encourage performance, risk-taking and entrepreneurial innovation. In particular, small and middle-sized businesses would receive preferential support.

Applied research in public institutions would focus on complex long-term research in transport, communications, aerospace and defence. A better balance would have to be found between the need to protect these areas from short-term political and economic influences and the con-

sequent dangers of bureaucracy and institutionalized decreases in effectiveness and creativity. In general, however, support would be for institutions rather than for projects, partly to reduce the administrative costs involved in assessment. Lenzer points out that the Ministry of Research and Technology employs more than 1,200 advisers at a cost of more than DM600 million. Many members of assessing committees come from large industries and have tended to channel funds away from smaller businesses.

With an eye to French competition, Heinz Riesenhuber, CDU/CSU spokesman on energy, stresses the role of nuclear power in the West German energy programme. CDU/CSU would be more aggressive in the application of proven techniques and development of reprocessing facilities. Highly critical of the ambivalence and delays of the Social Democratic Party (SPD), CDU/CSU would increase state participation in innovative developments such as the prototype reactors, in spite of soaring costs (see *Nature* 26 August, p.783).

In microelectronics, besides software and systems technologies, emphasis would be on widening the technological basis and development and production techniques for highly integrated elements and instruments in communications. CDU/CSU recognize the potential of biotechnology and promise increased state support for work in the field.

On environmental problems, stress is laid on the necessity for providing early anticipatory legal frameworks and also on the possibility of reconciling protection of the environment and economic growth by, for example, investment in energy saving technologies.

A deteriorating economic situation has resulted in no growth in gross national product in the first half of 1982, an increase in unemployment of 49 per cent over 1981 and hard times in the universities and elsewhere.

Sarah Tooze

scientific and technical meetings. The committee would not include outside academics. "We want to be sure that the US taxpayer benefits from this research and that the Soviet Union doesn't", says one official.

The academic community has been predictably alarmed by the incident. Officials of the Association of American Universities, the National Association of State Universities and Land Grant Colleges and the American Council on Education raised the matter at a meeting on 1 September

with Under-Secretary of State Richard D. De Lauer, the Pentagon's chief scientist. The Defense Science Board, which advises De Lauer, has made a series of studies on which scientific information should be withheld and has helped to compile the militarily critical technologies list, which is considered so sensitive that even the list of headings remains classified. Thus, so far, scientists and engineers have no idea what subjects the Department of Defense considers too sensitive for presentation in public.

Deborah Shapley

Keyworth on fusion

US seeks more foreign links

Washington

The Reagan Administration's alarm about the outflow of scientific information from the United States seems not to have interfered with its little-noticed plan to promote international cooperation in fusion. The architect of the policy is George A. Keyworth III, the presidential science adviser, who says that the next generation of fusion devices should be jointly agreed by the United States, the European Community and Japan, perhaps even jointly funded and operated by them.

Fusion advocates in Congress think the strategy too risky, citing the strains that might arise if construction of the \$1,000 million-dollar devices become subject to other strains among the allies. Another problem is whether the Administration would agree to the participation of the Soviet Union. The Soviet Union has an advanced fusion programme and a long history of cooperation with US fusion scientists. But President Reagan has been allowing bilateral science agreements with the Soviet Union to expire, and the agreement which covers fusion cooperation is due to end next March.

The White House's logic was explained recently by John Marcum, assistant director of Keyworth's office for energy and natural resources. "We are spending nearly half a billion dollars a year on fusion, which is the most challenging project man has ever undertaken including the Moon landing and the Manhattan project", says Marcum. "Yet we don't expect fusion to generate significant quantities of commercial electricity until the middle of the next century. Devices will cost in the billion dollar range and several different prototypes will have to be built before we find the one that is most practical."

"Let's face it, the United States, Japan and Europe can't afford independently to build their own tokamaks, stellerators and tandem mirror devices."

The JT-60 tokamak being built in Japan is now the most expensive fusion device under construction; it will cost an estimated \$1,000 million. The Joint European Torus (JET) that the European Community is building at Culham in the

United Kingdom will cost approximately \$500 million. The largest US device under construction is the TFTR at Princeton University, which will cost approximately \$300 million. Another device, the long-pulse length Torsupra being built by France, will also be expensive. Keyworth and others in the fusion community argue that by the time the next generation comes to be designed the costs will be too great for any single country to bear.

Keyworth has reportedly visited the most advanced example of US cooperation in fusion research, the US-Japanese Doublet III, built by the General Atomic Company in San Diego, California, and funded and operated on a 50:50 basis. The United States also participated with the EEC, Japan and the Soviet Union in an International Atomic Energy Agency design workshop known as INTOR (for International Tokamak Reactor).

As a result of the Versailles summit meeting in June, Keyworth has a chance to make his influence felt. He is the President's representative to the summit working group on cooperation in science and technology, and fusion is one area the group is studying.

The Department of Energy (DoE), which runs the US fusion programme, is following the Keyworth policy. Last week,

fusion programme officials met representatives of three foreign programmes in Washington: Donato Palumbo, the director of the EEC's fusion programme; S. Mori, a director of the Japanese Atomic Energy Research Institute; and R.S. Pease, programme director of the UK Atomic Energy Authority fusion programme. The purpose was to "look for areas of mutual technical interest" according to Mike Roberts of DoE. The foreign fusion officials also met officials of the White House, since they were in Battimore attending a world fusion conference.

But Congresswoman Marilyn Lloyd Bouquard (Democrat, Tennessee), who hosted a workshop last week where international fusion scientists met congressmen and their staffs, is sceptical of Keyworth's enthusiasm for international cooperation on future devices. One of her staff pointed out that it took two years of negotiations to get US-Japanese agreement on the Rotating Target Neutron Source machine at Lawrence Livermore Laboratory. "If you don't want to spend money, it's nice to throw this idea out", he went on. "On the other hand, there's more to international cooperation than good intentions. When we have some friction with our allies over steel or the pipeline, its going to be tough to go on building billion dollar machines."

It is not yet clear whether the Soviet Union would be deliberately excluded from future planning to decide which country builds a stellerator, or a mirror, or a tokamak. Keyworth's office makes it clear that it wants to enhance cooperation in EEC and Japanese projects, rather than in the International Atomic Energy Agency's joint project which includes the Soviet Union. Although Keyworth has not said publicly that he would exclude the Soviet Union, the Administration's overall denial of advanced technology to that country would make such a policy likely. It remains to be seen what Keyworth's would-be partners in fusion cooperation think on the question of Soviet participation.

Deborah Shapley

University teachers take to the streets

Politicians in Britain are well aware that further cuts in spending on the universities will be met with indifference or approval from the majority of their constituents. That, at least, is the thinking behind a new advertising campaign launched by the Association of University Teachers (AUT) last week. To give added weight to the AUT's political lobbying the association is buying advertising space on buses on London's 88 route, which passes the AUT's headquarters and, as it happens, goes down Whitehall. AUT sees this as a beginning to the process of making the general public more aware of the value of the universities. Further exploits being considered include advertising on London's underground trains and spreading the bus campaign to provincial cities.



Clinch River fast breeder

Now the crunch

Washington

Opponents of the Clinch River Breeder Reactor project in the US Congress seem well placed this year to bring the project to an end. Ever since the Clinch River project began a decade ago, scarcely a year has passed without an attempt to kill it. President Carter tried, congressional committees have repeatedly tried and last year the House fell just 20 votes short — and the Senate just two votes short — of withdrawing funds.

This year congressional opponents are trying again and have put together an unusual alliance of liberals and conservatives that seems likely to prevail. The rallying point for this alliance is economics, an issue that the breeder's opponents believe has been strengthened by several developments:

- The breeder's competitiveness. A study by Los Alamos National Laboratory concluded last year that uranium would have to rise from its present price of about \$25 per pound to \$165 per pound before the breeder would break even. According to the Department of Energy (DoE), the price will have reached only half that break-even point by the year 2020.

- Cost overruns. Clinch River was initially supposed to cost \$400 million but the official projection now is \$3,570 million. In addition, all of the costs overruns are to be borne by the government, while industry's contribution is fixed at a maximum of \$257 million. Thus the project, which was supposed to be an industry-government partnership, is now government-supported to the tune of more than 90 per cent.

- Cheaper alternatives. The efficiency of conventional light-water reactors can be increased by 20 to 25 per cent through modifications to allow so-called "extended burn-up" of fuel. Breeder opponents say this is a far more cost-effective way to stretch existing supplies of uranium. Both Westinghouse and General Electric are already collaborating with Japanese firms on the design of these advanced reactors.

- Budget cuts and tax increases. Many members of Congress — including Republicans — are finding it difficult to justify expenditures on the breeder in the face of massive budget cuts in social programmes and the recent tax increase.

Among the most visible opponents this year are such staunchly conservative groups as the Heritage Foundation, which find the Reagan Administration's support for Clinch River inconsistent with its often-stated free-market policy. Even strongly pro-nuclear legislators are coming out against it; Senator Gordon Humphrey (Republican, New Hampshire) recently called it "nothing but a plutonium pork-barrel", a reference to the growing conviction that the Administration's support for

the project is a political favour to Senator Howard Baker, the Republican majority leader from Tennessee, in whose state the Clinch River site lies.

The showdown will come in Congress before 30 September, the end of the fiscal year. Any appropriations bills that are not passed by that date will be incorporated into what is known as a "continuing resolution", a stop-gap measure that continues spending at current levels. So far, not a single appropriating bill has been acted on; and it is thus very likely that the Clinch River appropriations will end up in such a continuing resolution.

That is exactly what Clinch River's supporters want. Continuing resolutions are usually not subject to amendment, but are presented to the Congress as a take-it-or-leave-it package. By incorporating appropriations for Clinch River into a bill that includes many vital appropriations, it is much more likely to go through.

Opponents, led by Representative Claudine Schneider (Republican, Rhode Island), have already anticipated that and are seeking a special rule that would allow a separate vote on the House floor on the \$252 million request for Clinch River. (The remaining \$500 million or so budgeted for the breeder programme is not at issue. Those funds are for research and for the Fast Flux Test Facility, a small experimental breeder at the Hanford Reservation in Washington state.)

Meanwhile, supporters of Clinch River suffered another blow recently when a federal district judge in Atlanta issued an injunction preventing initial site preparation for the project. The order had been sought by the Natural Resources Defense Council, an environmental group, which charged that the Environmental Protection Agency broke the law by allowing construction to begin without an environmental impact statement and a water permit. The DoE, which is managing the project, had been eager to start on the actual construction before Congress took up the matter.

Still, supporters maintain that it makes no sense to turn back now. According to the American Nuclear Energy Council (ANEC), an industry lobbying group, research and development on Clinch River is 96 per cent complete while 60 per cent of the plant's major components are either completed or on order. Industry is also threatening to demand a refund on the \$122 million it has already contributed should the project be terminated.

The supporters have also sought to make Clinch River a symbol of nuclear power and American technical superiority in general. Yet even nuclear "hawks" are far from enthusiastic about Clinch River. And the "keeping up with the Joneses" (in this case France) argument is piling beside the economics. France's Super Phénix breeder, says one Clinch River opponent, has become the "nuclear Concorde".

Stephen Budiansky

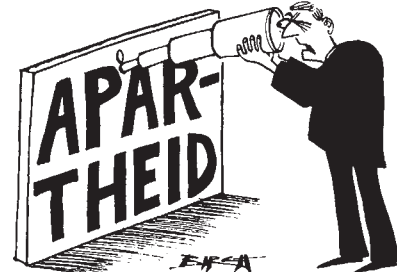
SERC in South Africa

British astronomers seeking to sever links with the South African Observatory near Sutherland, Cape Province, seized another opportunity to put their case at a symposium arranged by the Science and Engineering Research Council (SERC) two weeks ago. But the balance of opinion still seems to rest with those who claim that the link with South Africa is scientifically invaluable.

At the symposium, held at the University of Sussex and arranged to assess future commitments to the Anglo-Australian Telescope (AAT) and the South African Observatory (SAO), Dr Michael Rowan-Robinson of Queen Mary College, London, argued that if British astronomers did not face up to the moral and political problems occasioned by continued collaboration with SAO, they might find the decision taken out of their hands.

At present SERC is a major user of the observatory, buying 35 per cent of the total viewing time at an annual cost of about £240,000. SAO is a valuable

"I STILL CAN'T MAKE IT OUT..."



facility for British astronomers because of its geographical and climatic position; most of the work done there could not be reallocated to AAT.

The only feasible alternative facility for the optical and infrared photometry work largely done at SAO would be the European Southern Observatory in Chile. But, as Professor Michael Disney, one of the members of the SERC working party which has been examining the case of renewing the contract with the SAO, points out, this alternative may well prove more politically unsavoury to astronomers, as well as more expensive.

When the SERC working party questioned the groups likely to want to use SAO, however, it received only four objections on political grounds to its continued use, whereas 43 other groups wanted to carry on working there. So, in spite of the protests of Dr Rowan-Robinson and others who argued at the symposium in favour of abandoning SAO, it seems likely that the Astronomy, Space and Radio Board, when it meets in December, will make its decision primarily on the scientific merits of SAO.

David Millar

French budget

Science pluses

Paris

The French science budget really is on the up and up: FF32,500 million (£2,650 million) in 1983, 28 per cent more in 1982, according to figures released last week by the ministry of research and industry.

This figure, which includes all government civil research and development, must be corrected for inflation (now 12 per cent and optimistically projected to be 8 per cent next year). But even so the increase is impressive and seems even more remarkable when set against the present parlous state of the French economy, propped up last week by a \$4,000 million international loan, and against the average government budget increase for next year — a meagre 5 per cent in real terms.

How does Jean-Pierre Chevènement, the minister for research and industry, do it? From the left of the socialist party now in power, he appears to have formed a remarkable liaison with a man from the right, Jacques Delors, the finance minister, who is constantly urging budgetary restraint and the control of inflation. Both, it seems, are seized with a kind of constructive realism: they agree that France must renew and rebuild its wealth-generating industry, to which end science and technology are essential.

This is an economic objective, however, not a "scientific" one, and its implications run right through the 1983 budget. Moreover, Chevènement has stealthily adopted, without ever quite announcing it, a means of science funding akin to the British "Rothschild principle": the customer pays, and the contractor does what the customer wants, to the best of his ability. In France this is now called budget management "*par programme*".

The ministry identified seven major strategic areas, and four key technologies (called "*programmes mobilisateurs*" and "*programmes de développement technologique*", see below). The only difference from Rothschild's historic adjustment to the accounting of British research council cash in 1972 is that, in France, no precise figure has been set aside for the research councils to manage by themselves.

The seven *programme mobilisateurs* are:

- Rational use of energy and diversification of sources
- Biotechnology
- Development of the electronics industry
- Science and technology for the Third World
- Research on employment and work conditions.
- Promotion of French as a scientific language, and the spread of scientific culture
- The penetration of technology into industry

The four *programmes de développement technologique* are: Nuclear power; Space; Civil aeronautics; Oceans.

The whole sum, effectively, is in the hands of the customer — the ministry.

This shows up particularly clearly if the FF32,500 million civil research budget is divided into two parts: money directly controlled by the ministries or sent to industry, and money "processed" through the *grands organismes* (on which French university and *grands organismes* researchers can exert the greatest pressure). The first category is more applied, and more easily controlled politically; the second more basic, and less easily controlled.

The easily controlled category includes: FF1,500 million directly available to Chevènement for his own research priorities (up 29 per cent on this year); FF1,800 million for industrial research support (up 78 per cent); and FF7,200 million spent on research and development by other ministries under Chevènement's guidance (up to 65 per cent). Altogether, this category totals FF10,500 million in the 1983 budget, compared with only FF6,500 million in 1982: a rise of 62 per cent. The remainder of the budget, sullied — from the political point of view — by the inertia of the *grands organismes*, amounts to FF22,000 million, a rise of only 17 per cent on this year.

These global figures, however, conceal important differences in treatment among the different *grands organismes*. Here it is also important to distinguish between money credited — actually to be spent next year — and authorized, which equals the credits plus sums, usually for capital equipment, which can be committed but not spent. A third accounting category is "ordinary spending", essentially salaries. All the figures so far given apply to the authorized sums plus ordinary spending, as this is the best long-term indicator.

In the 1983 budget, the most dramatic differences are in the credits, which for the *grands organismes* rise 26 per cent over the last year to FF8,900 million. (The fact that authorizations rise only 16 per cent indicates perhaps that the ministry is taking care not to over-commit itself to basic research for 1984, should the French budget collapse entirely.)

Within these credits to the *grands organismes*, the Centre National de la Recherche Scientifique (CNRS), which does most basic research and has the greatest prestige, will see a 28 per cent rise in 1983 and — almost uniquely, a larger rise in authorizations: 36 per cent. CNRS will do well. INSERM, which supports medical research, gets a rise of 37 per cent in credits, but only 21 per cent in authorizations.

This year, Chevènement has been able to amass enough cash essentially to insulate basic scientists from the impact of technological priorities, which are appearing as additions to budgets rather than alterations. But whether this will be so next year (that is, in 1984) may depend on the performance of the French economy in the interim.

Robert Walgate

Council of scientific unions

Enter China

The admission of the People's Republic of China to the International Council of Scientific Unions (ICSU) was the highlight of this year's general conference, held last week in Cambridge, England. China will be represented by the China Association for Science and Technology (CAST).

During the past few years, science — a major casualty of the cultural revolution — has been considerably rehabilitated in China. The prestigious journal *Scientia Sinica* has resumed publication and the admission of China to ICSU was so little in doubt that a seven-person delegate from Beijing was already in Cambridge, waiting for two days while a suitable formula was worked out. The main problem was to ensure the admission of CAST without excluding the academy in Taipei (Taiwan), whose contribution to science has of recent year been considerably larger than the modest size of the population would suggest.

Apart from excitement over the Chinese issue, which some delegates felt had become too political, this ICSU conference, like its predecessors, consisted largely of reports on continuing activities rather than startling innovations. There were no real controversies, even on such sensitive subjects as the storage of high-level radioactive wastes on or near the Earth's surface; the main controversy was whether they should be so stored for 50 or 100 years. There was, however, general agreement that an ICSU-sponsored survey of nuclear war scenarios would be useful.

In the Committee on the Safeguarding of the Pursuit of Science, there were some sharp remarks on recent cases of Soviet Jews deprived of their academic degrees for political reasons. On the other hand, the case reported to the Committee for the Free Circulation of Scientists of the two Soviet scientists refused visas for an international conference in Australia in August was finally resolved, after an exchange of telexes, by a pledge from the Australians that no such difficulties would arise in the future for Soviet scientists wishing to attend multinational conferences in Australia.

A number of routine matters were brought to a satisfactory conclusion. The Committee on Science and Technology in Developing Countries (COSTED) was reorganized to give it a wider orientation (recently it has concentrated almost entirely on India). A long-term plan by the Committee on Publications and Communications (COPAC) for the establishment of an ICSU press was accepted, and COPAC was authorized to proceed to the projected first phase of that plan — the discussion of ways and means with the presses of member national academies and the like, as well as with commercial presses where academies do not have their own press.

Vera Rich

Biotechnology

Insulin on tap

Insulin has become the first product of genetically manipulated bacteria to be marketed for human use. Now that Eli Lilly has official approval for sale in the United Kingdom of bacterially-produced human insulin, it will try to wrest from Novo Industry some of the £20 million annual market for insulin in the United Kingdom.

Approval has come one month after publication in *The Lancet* of a double-blind clinical trial involving five UK hospitals and 94 insulin-dependent diabetics. The trial demonstrated only slight differences between Lilly's new insulin, which has the exact structure of the human hormone, and conventional preparations purified from pig or cow pancreas. Probably because it is absorbed slightly more rapidly from the site of injection, rather more of the new insulin was needed to produce the same reduction of blood sugar. That slight disadvantage, however, is offset by corresponding advantages, according to Professor Harry Keen of Guy's Hospital Medical College in London.

The possible advantages are two-fold. First, the few diabetics who develop an allergy to animal insulins may not do so to human insulin. And second, the new insulin is slightly better than the old at hastening the removal of ketones from the bloodstream.

Availability and cost, therefore, will determine the choice of insulin in the market. On the basis of a 1978 report of the US National Diabetes Advisory Board, Eli Lilly believes that the demand for insulin will outstrip conventional supplies before the end of the century, mainly because of the likely increase in insulin-dependent diabetes.

As regards cost, Eli Lilly claims to be undercutting by about 10 per cent the cost of Novo Industry's "humanized" insulin. Novo, which dominates the European insulin market, recently perfected the trick of chemically converting purified pig insulin into human insulin. Its "humanized" insulin beat Lilly's human insulin to the marketplace when it received UK approval last June.

In an attempt to beat off the opposition from Lilly, Novo had contemplated withdrawing all its conventional animal insulins, thereby forcing a switch to its human version. If that had happened, one reason for not switching to the Lilly insulin — that it was human in structure — would have been removed before it arose. However, Novo's mooted switch would have meant an increase in price, and for that reason at least the idea was dropped.

Eli Lilly will now try to muscle in on the UK insulin market with its new product, having spent a great deal of money on softening up the medical consumers by

ostentatious image advertising in medical journals before approval was granted. It is less clear what will happen in the United States where Lilly has about 85 per cent of the market already. A spokesman for Lilly said he could not comment on the question of whether approval for its new insulin has yet been sought in the United States.

It has taken just four years for Lilly to scale up the process developed by Genentech for the production of human insulin from two bacterially-produced fragments. The bacteria are supplied with synthetic genes for either the A or the B fragment of insulin. After the two fragments are produced by the bacteria at Lilly's Indianapolis site they are chemically combined either there or at the site in Speke, near Liverpool. Final purification is carried out at Speke. **Peter Newmark**

Indian space effort

Falling flat

New Delhi

from a correspondent

India's grandiose scheme for satellite-based domestic communications and direct television broadcasts has received a jolt following the sudden demise of the INSAT-1A satellite. The \$60 million multi-purpose satellite, which had a design life of seven years, was abandoned on 5 September after only four months in space. The Indian Space Research Organization (ISRO), which was responsible for its operation, said the satellite had become useless because it had no more fuel left for on-orbit control.

Ever since its launch from Cape Canaveral on 10 April, INSAT-1A had been plagued by a series of problems. For several days, the INSAT control team at Hassan near Bangalore struggled with the jammed C-band antenna and a stuck solar sail. The antenna was deployed after much expenditure of fuel but the team had no luck with the solar sail, an umbrella-shaped device meant to offset the imbalance caused by INSAT-1A's asymmetrical solar panels. In the absence of the solar sail, on-board thrusters were fired more frequently to keep the satellite balanced, so depleting the fuel. ISRO estimated in mid-June that the fuel would last until the middle of 1984, but this calculation was upset by yet another problem detected early in September. The isolation valve in the fuel manifold got stuck and the fuel simply vented into space, leaving ISRO with no alternative but to abandon the satellite.

The satellite had been partly operational since 15 June from its geostationary parking slot at 74°E, although its capacity was largely under-utilized because of problems on board the spacecraft and non-readiness of the ground segment. INSAT-1A was used to provide 200 long-distance telephone circuits (against its capacity of 4,000) and television relaying was restricted to four hours a day because of overheating

of the television transponder on board.

The loss of INSAT-1A has placed the Indian government in a difficult position. Apart from the monetary loss (part of which is covered by insurance) the country faces the prospect of keeping idle the 31 Earth stations set up at a cost of more than \$60 million. The enormous preparations for live television broadcasts of the Asian Games scheduled for November have also been jeopardized by the failure of INSAT-1A. While the radio and television networking should await the launching of INSAT-1B, the impact on the Indian Meteorological Department is minimal as it continues to get cloud pictures and cyclone warnings from the United States weather satellites.

To restore a semblance of domestic communications and to lessen the impact of INSAT-1A's failure on the Asian Games coverage, India is planning to lease two transponders of Intelsat's Indian Ocean satellite at a cost of about \$16 million a year.

NMR tomography



One of the latest images produced by the rapidly developing technique of nuclear magnetic resonance (NMR) tomography. This image was obtained in the Erlangen Laboratories of the German company Siemens AG, which hopes to market a machine for hospital use early next year, some months after the first machines are expected to be sold by Picker International. The images reflect the precessing movement and dipole characteristics of protons subjected to a high frequency magnetic field. The relaxation time of protons so affected depends upon their environment. Different environments are imaged as different shades of grey. Because it detects protons, NMR tomography provides a different picture from that of the X-ray tomography used in current brain and body scanners. Furthermore, there are no known hazards of magnetic fields to rival those of X-rays. On the other hand, the resolution of NMR tomography is inherently less good than that of the X-ray scanners (and it will wipe out credit cards — including those of potential customers with about £0.5 million to spare).

CORRESPONDENCE

Pitdown jaw confirmed as orang

SIR — When the specimens from Pitdown were “discovered” in 1912, the combination of the human skull with an ape-like jaw caused immediate controversy^{1,2}. Smith Woodward, G.E. Smith, A. Keith and others accepted the association, but G. Miller (1915) thought that the jaw represented a new species of chimpanzee, *Pan vetus*³. Hooton accepted the association but described the jaw as “almost indistinguishable from that of a young

the canine tooth into the gravel pit must have known that the jaw was also that of an orangutan.

The positive identification of the Pitdown jaw and tooth as orangutan, after 70 years of often acrimonious debate, is of more than academic interest. Although the Pitdown “fossil” was a forgery, many primate fossils have been broken and largely destroyed by natural causes. The relations of human beings

Collagen reactions: Pitdown jaw and canine

	Human	C. chimp	P. chimp	Orang	Rhesus	Elephant	Bovine
Jaw	0.61	0.56	0.76	1	0.35	0.01	0.04
Canine	0.66	0.62	0.42	1	0.34	0.02	0.04

The numbers represent relative binding of antisera to various collagen species by extracts of Pitdown jaw and canine. Maximum binding is taken to be 1. Antiserum to orang collagen is more strongly bound by the Pitdown extracts than antisera to any of the other collagens. The pattern of cross-reactions is also characteristic for orangutan collagen.

chimpanzee”. There were doubters, however, and Weidenreich (1946) believed that the jaw did not belong with the skull and that it was the jaw of a fossil orangutan⁴. Weidenreich spoke of *Eoanthropus* as the result of reconstruction by “English authors”. Clearly, distinguished scholars could make the facts fit any of the suggested theories. In retrospect, it is surprising that the fragmentary mandible could be diagnosed at all.

Overconfidence in comparative anatomy led prominent scientists astray. Weiner (1955) describes the story of the discovery of the forgery¹. It was chemistry, not anatomy, which led to the proof that the jaw and skull did not belong together⁵. Weiner¹ makes a strong case that the jaw and canine tooth are those of an orangutan, but he must resort to comparisons of the same kind that led Keith, Hooton and many others to conclude that they were like those of a chimpanzee. We report here the results of a collagen radioimmunoassay which proves that both the Pitdown jaw and the canine tooth are those of an orangutan.

Collagen radioimmunoassay¹ permits identification of species-specific proteins in very small samples of tissue. The method has been used to investigate the relations of fossil to living species^{6,7}. Collagen identification was performed using 50 mg of bone from the Pitdown jaw and 0.9 mg from the canine tooth. Antisera to orangutan collagen were bound by extracts from the jaw and tooth more strongly than antisera to the collagens of man (*Homo sapiens*), common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (*Pan paniscus*), rhesus monkey (*Macaca mulatta*), Indian elephant (*Elephas maximus*) or cow (*Bos taurus*), as shown in the table.

These results confirm that the jaw is that of an orangutan and that it did not belong with the human skull. The canine tooth was found a year after the jaw, and it has been suggested that it was introduced into the gravel by a second forger. But at the time of the “discoveries”, everyone was stressing the chimpanzee-like characteristics of the jaw. It was more than 15 years before the similarities with the orangutan were noticed. Unless one assumes, as Matthews⁸ does, coincidental use of orangutan material by two different forgers, it would seem likely that whoever put

to other primates are still debated, and neither anatomy nor currently available fossils are likely to settle the problem of “man’s place in Nature”. Whenever possible, such problems should be settled, or at least bounded, by the findings of molecular biology.

JEROLD M. LOWENSTEIN

Department of Medicine,
University of California,
San Francisco, California, USA

THEYA MOLLESON

British Museum (Natural History),
London SW7, UK

SHERWOOD L. WASHBURN

Department of Anthropology,
University of California,
Berkeley, California, USA

1. Weiner, J.S. *The Pitdown Forgery* (Oxford University Press, 1955).
2. Reader, J. *Missing Links* (Little Brown, Boston, 1981).
3. Miller, G.S. *Smithsonian misc. Collns* **65**, 1-31 (1915).
4. Weidenreich, F. *Apes, Giants and Man* (University of Chicago Press, 1946).
5. Weiner, J.S., Oakley, K.P. & Le Gros Clark, W.E. *Bull. Br. Mus. nat.Hist.* **A2**, 141-146 (1953).
6. Lowenstein, J.M. *Phil. Trans. R. Soc. B292*, 143-149 (1981).
7. Lowenstein, J.M., Sarich, V.M. & Richardson B.J. *Nature* **291**, 409-411 (1981).
8. Matthews, L. H. *New Scientist*, **90-91** (series of 10 articles on Pitdown; 30 April-2 July 1981).

Ball of fire

SIR — I was interested to read Professor Pippard’s letter on his experience with lightning (*Nature* 19 August, p.702). Thunderstorms are frequent in the Entebbe Peninsula, Lake Victoria (Uganda). During one of these storms, which usually come at night time, there was a simultaneous flash of lightning and its associated clattering crash of thunder. A second or less later, several balls of brilliant blue light, about 4-6 cm diameter, entered the room through a window on the south side and “floated” across the room to leave by a window on the east side. My wife and I were already awake (it would have been difficult not to be) and independently exclaimed aloud on what we had just seen.

The incident occurred during the pre-independence period of Uganda and so there was no air conditioning. We therefore left all the windows open at night, protected only by the anti-mosquito metal screening. During many hundreds of storms we had never seen

this phenomenon before, even when the house was struck, as it was on several occasions. We therefore continued with our practice of leaving all the windows open. But during the same rainy season the same thing happened again with the same startling effect. After that we always kept the windows on one side of the room closed and we never experienced the phenomenon again.

J.D. GILLET

London School of Hygiene and Tropical
Medicine, London WC1, UK

Monstrous outrage

SIR — Since it appears to be the season for arguing about names, permit us to point out that Jeremy Greenwood¹ does not really appear to have a valid case for complaint under the International Rules for Zoological Nomenclature when a contributor to *Science*² uses the name *Prion*, otherwise applied to a group of birds, to Proteinaceous Infective Particles, since these can clearly only be classified as animals when they appear in them as inclusions. *Nature* itself has in fact committed a much more serious nomenclatorial offence in the past by publishing a new name *Nessiteras rhombopterox* for the Loch Ness monster⁴, when *Nessum monstrosum* had at least 40 years’ priority⁴, unless, as Gregory’s pictures suggest⁴, there is more than one type of monster in the loch. In the present case would it not be simpler just to call the particles “pips”, or would this lead to further complaints from botanists?

W. R. BOURNE

Department of Zoology,
University of Aberdeen

S. HOWE

BP Petroleum Development Ltd,
Aberdeen, UK

1. Greenwood, J.J.D. *Nature* **298**, 510 (1982).
2. Prusiner, S.B. *Science* **216**, 136 (1982).
3. Scott, P. & Rines, R. *Nature* **258**, 466-468 (1975).
4. Gregory, W.K. *Nat. Hist.* **89**, 130-131 (1980, reprinted from 1934).

Which war?

SIR — There is a minor error in your editorial “Rules for limited war” (*Nature* 27 May, p.253). Surely you mean that it was the 1967 “six day war” and not the 1973 “Yom Kippur war” in which the Israelis took the Sinai Desert. The Yom Kippur war was the one in which the Israelis, already in possession of the Sinai, were the victims of a surprise attack that resulted in no real net gain of territory for anyone. Nonetheless, thank you for a timely and interesting editorial.

BERT ATMSA

Montclair, New Jersey, USA

French Nobels

SIR — In the article “Language of love” (more like *The Sun* than *Nature*, but let it pass) in *Nature* of 26 August (p.780) you describe Alfred Kastler as “France’s sole Nobel physicist”. May I remind you that Louis de Broglie, the discoverer of wave mechanics and 1929 Nobel Laureate is still with us and that Louis Néel got the Nobel in 1970 for his discovery of antiferromagnetism.

ANATOLE ABRAGAM

Collège de France, Gif-sur-Yvette, France

NEWS AND VIEWS

Proton decays and high-energy speculation

from Louis Lyons

A HIGHLIGHT of the recent International Conference on High Energy Physics* was the suggestion that catalysed decay of protons might occur. It has long been realized that protons must be extremely stable. As Maurice Goldhaber has said: "You can feel it in your bones that protons do not decay"; for, if their lifetime was shorter than 10^{17} years, the damage to bones would be serious. Recent grand unified theories (which unite the weak force responsible for β decay with the strong one which confines quarks within protons) suggest, however, that protons could decay with a lifetime of around 10^{31} years. Such a value would give a reasonable chance of the decay of just one proton in a human being in a lifetime. Experiments looking for proton decays, necessarily involving hundreds of tons of material situated deep underground, are currently being run.

The new suggestion, originally put forward by V. Rubakov (Institute for Nuclear Research, Moscow) and independently developed by C. Callan (Princeton University) is that, if magnetic monopoles exist, they could, in their passage through matter, fairly readily stimulate proton decays. Grand unified theories do suggest that magnetic monopoles exist, although the number near the Earth and at this stage of the Universe's development is unknown.

In passing through a large detector, a monopole could produce a string of proton decays, separated from each other by perhaps tens of centimetres. Such a sequence of events correlated in time to within a few microseconds would be a very characteristic experimental signature, and has provided an extra motivation for the proton decay experiments. It has also led to the somewhat tongue-in-cheek suggestion that if only such monopoles could be collected, the ensuring catalysed proton decays may provide a source of energy.

Such exciting chains of decays have not yet been observed. Present experiments

merely set limits on the product of the flux of monopoles times the probability of their catalysing proton decay. The observation of decays would be a major development, shedding light both on the existence of monopoles and on the instability of the proton.

Meanwhile, experiments are accumulating events which could contain examples of the ordinary decay of the proton. The main difficulty is that the expected decay rate is so low that all sorts of background effects can be confused with genuine decays. An experiment at the Kolar Goldfield in India has three candidates, while another inside Mont Blanc has one. More data and further studies of backgrounds are now required. If the results are confirmed, it would open a whole new field for experimental study, and would be a major success for the grand unified theories which predicted the decays.

Four groups working at the new pp collider reported their results at the conference. The supersynchrotron (SPS) at CERN in Geneva is used to accelerate simultaneously protons and antiprotons circulating in opposite directions around the machine. When both particles reach 270 GeV, they are brought into head-on collision, resulting in interactions at the highest artificially produced energies. Because of relativistic effects, head-on collisions are very much more effective than those of a single beam with a stationary target. With the extra energy available, hitherto unobserved phenomena could take place. Until now, however, the low intensity of the beams has resulted in only a few thousand interactions being observed, so that only the gross features of the reactions have been studied. Thus, although the heavy W and Z bosons responsible for transmitting the weak interactions (in much the same way as the photon transmits electromagnetic phenomena) can be produced in $\bar{p}p$ reactions, they have yet to be observed.

Although free quarks have not been produced in experiments at accelerators (and perhaps have not been seen elsewhere — see L. Lyons *Nature* 291, 534; 1982), the

existence of quarks has a readily apparent effect on many types of high-energy reactions. Thus, in electron-positron annihilations, the various types of quarks contribute once the beam energies are high enough. Data presented at the conference provided new information concerning the properties of the bottom quark (*b*), especially concerning its decays to lighter quarks. Searches for evidence of the expected top quark (*t*) have still been unfruitful.

High-mass particles usually decay faster than light ones, so techniques for detecting tracks corresponding to such particles and for measuring their lifetimes need continual improvement. Whereas the weakly decaying strange particles have lifetimes down to 10^{-10} seconds, the heavier charmed particles live for less than 10^{-12} seconds. This corresponds to a distance of 2 mm for a charmed particle travelling at 99 per cent of the speed of light. The lifetime of the τ particle, the heavier version of the electron and the muon, is also in the latter range. The conference saw a series of measurements of these lifetimes, which now begin to look reliable, with the systematic effects better controlled. An interesting question for the future is whether the lifetimes of particles incorporating the *b* quark will be similarly amenable to direct measurement, after further improvements in the techniques.

Another impressive result came from the measurements of the forward-backward asymmetries in the reactions $e^+e^- \rightarrow e^+e^-$, $\mu^+\mu^-$ and $\tau^+\tau^-$ at high energies. The main contribution to these processes comes from a virtual-photon intermediate state, which on its own would result in a symmetric angular distribution. However, its interference with virtual weak boson Z^0 results in an expected asymmetry, whose magnitude is consistent with observations. This, together with the observation of parity violation in an atomic physics experiment, provides further confirmation

*The XII International Conference on High Energy Physics was held in Paris, in July 1982. Some 1,200 theorists and experimenters from 45 countries gathered to hear over 100 talks.

Louis Lyons is in the Nuclear Physics Laboratory, University of Oxford, Keble Road, Oxford OX1 3RM.

of Weinberg and Salam's model of electromagnetic and weak interactions.

As usual, theoretical speculation abounded. One relatively new approach is the idea of supersymmetry, relating $\frac{1}{2}$ -integer spin fermions and integer spin bosons. Experiments are already setting limits on the possible existence of the new objects which are expected in such models, although in most cases these are, as yet, not particularly stringent. Another series of models deals with the possible non-elementary nature of quarks and leptons (see L. Lyons *Nature* **298**, 601; 1982).

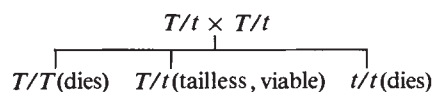
By and large the rest of the conference was devoted to clarification of experimental issues. Each on its own may seem, at the time, not particularly significant, but the increase in our appreciation (if not understanding) of elementary particle phenomena from one conference to the next is impressive. Certainly the prospects of new results to come from the proton decay experiments and from the $\bar{p}p$ collider make it look as if the next couple of years could bring excitement and surprises in an already fast-moving field.

Analysing the mouse *T/t* complex

from P. W. Andrews and P. N. Goodfellow

RECENT experiments, reported by Artzt and colleagues in *Cell*^{1,2}, show that progress is being made in the detailed genetic analysis of the organization of the mouse *T/t* gene complex. Ever since the dominant mutation *Brachyury*, or *T*, was first described³ in 1927, the *T/t* complex has been a puzzle for geneticists and developmental biologists. The *T* mutation, located on chromosome 17, about 15 cM from the *H-2* region, is of itself unremarkable: it behaves as a single mutation that obeys the laws of Mendel, producing a shortened tail in heterozygotes and causing death of homozygotes after about 10 days of embryonic development. It is when *T* mutants are outcrossed to laboratory and wild mice that the strange properties of the mutation become apparent.

With a surprising frequency many such crosses yield offspring that completely lack a tail, and when intercrossed breed true, producing only tailless offspring⁴. The tailless condition is caused by interaction of *T* with a *t* haplotype; each *t* haplotype consists of a series of chromosome 17-linked genetic alterations which are responsible for a bizarre range of phenotypes. The *t* haplotypes isolated from wild mice carry recessive lethal mutations which complement the lethal effect of *T* so that intercrossing *T/t* (tailless) heterozygotes constitutes a balanced lethal system⁵:



Many such *t* haplotypes have been isolated and their lethal factors have been found to fall into several different complementation groups, each of which causes death at a particular stage of embryonic development^{6,7}. A number of other peculiar genetic traits are often exhibited by the *t* haplotypes, including segregation distortion in males but not females (*t*/+ heterozygous males often transmit the *t* and '+' chromosomes in non-mendelian proportions), sterility in males but not

females doubly heterozygous for two complementing *t* haplotypes, and suppression of crossing-over in the interval between the centromere and *H-2* on chromosome 17 (see refs 8–10 for general reviews of the *T/t* complex). This suppression of recombination is of particular practical significance for it has made the genetic dissection of the *T/t* gene complex extremely difficult.

Over the years, various proposals have been made to explain the properties of the *T* locus. Gluecksohn-Waelsch and Erickson^{8,11} suggested, by analogy with *H-2*, that genes of the *T/t* complex control the expression of cell-surface molecules that are important in mediating cellular interaction, in embryonic development and perhaps also during spermiogenesis and fertilization. Bennett and her colleagues have presented data indicating that the *T/t* complex does control the expression of cell-surface molecules on sperm and early embryos^{12–14}, possibly through the activity of glycosyl transferases¹⁵. However, others have failed to detect such specific antigens on sperm¹⁶ and data indicating that at least some of the *T/t* gene complex effects are associated with metabolic disturbances have been reported^{17,18}. An alternative view of the *T/t* complex is that it consists of a random selection of lethal mutations maintained in the population because of their association with a gene causing segregation distortion.

Despite suppression of crossing-over by *t* haplotypes, the occasional recombinant does occur in the *T* region when a *t* haplotype is present (about 0.2 per cent compared with about 7 per cent recombination between *T* and the coat morphology marker *tf* in the absence of a *t*-haplotype). Lyon and her colleagues used this phenomenon to study the genetic organization of

the *t*-haplotype *t*⁶ (refs 19–23). They isolated reciprocal recombinants and showed that the factor which interacts with *T* to produce the tailless condition (*t*⁷) is separable from the factor(s) which causes embryonic death (*L*) and male sterility (*S*). The factors causing segregation distortion were more difficult to analyse but seemed to result from the interaction of several different factors mapping to the regions carrying *t*⁷, *L* and *S*. Another factor, segregation distorter *A*, mapping separately from these was also detected.

A similar conclusion was reached by Styrna and Klein²⁴ who also noted a striking parallel with the genetics of transmission distortion and male sterility seen in the segregation distorter (*SD*) system of *Drosophila melanogaster*²⁵. In *Drosophila*, segregation distortion results from the interaction of two closely linked loci, *Rsp* and *Sd*, with effects on segregation and male fertility²⁶, that are very similar to those of the *t*-haplotype *A* and *S* factors. Many other parallels between the *T/t* complex and *SD* have been noted²⁷ and it is perhaps not a very big leap of imagination, given that the spermiogenesis must have originated early in phylogeny, to suppose that fundamental aspects of this process may be highly conserved and subject to similar genetic aberrations in such widely different species as *Drosophila* and mice.

A notable new advance in the analysis of the *T/t* complex was made a year ago with the report by Silver and Artzt that much higher levels of recombination occur between two complementary *t*-haplotypes than between a *t*-chromosome and a normal chromosome²⁸. This supports the idea, originally suggested by Lyon²⁹, that recombination suppression results from mismatching of *t*-haplotype and normal chromatin. Since all *t*-haplotypes are thought to have evolved from a common ancestor, the cause of such mismatching would be common to all *t*-haplotypes. Matching and recombination can occur between two *t*-chromosomes but not between a *t*-chromosome and a normal chromosome. This report opened the way for much more detailed genetic analyses of the *T/t* complex than had previously been possible.

Artzt and her colleagues have now studied recombination between four *t*-haplotypes, *t*^{w12}, *t*^{w32}, *t*^{w5} and *t*^{w8}. In the first of two articles¹, they provide evidence that the lethal factors of these *t*-haplotypes map to quite different positions and hence are not allelic. They also found in one case, *t*^{w32}, that embryonic lethality may be due not to one but to two interacting mutations at two different but closely linked loci.

One unexpected observation was the higher than normal recombination seen between *T* and *tf*. For example, between a *T tf t*^{w12} chromosome and *t*^{w5} or *t*^{w32} chromosome, 15 per cent recombination in the *T-tf* interval was recorded, compared with 10 per cent in the normal case (only

P. W. Andrews is in the Wistar Institute, Philadelphia, Pennsylvania 19104 and P. N. Goodfellow in the Imperial Cancer Research Fund Laboratories, Lincoln's Inn Fields, London WC2A 3PX.

females studied). In combination with t^{w18} , less recombination was measured (8.1 per cent) but this was interpreted in terms of a shorter region of t -chromatin in this t -haplotype — as suggested by the fact that recombination between tf and the t^{w18} lethal factor is not normally suppressed when t^{w18} is in *trans* to a normal chromosome 17. These abnormal recombination frequencies need to be viewed in the light of evidence for abnormal location of genetic loci in the t -haplotypes. The authors recall the suggestion that the locus, *Tcpl*, coding for the p63 protein, which is defined by two-dimensional gel electrophoresis, is differently located in t -haplotypes and the normal chromosome 17 (ref. 30). Furthermore, in their second article², they report evidence that the $H-2$ region in t -haplotypes is located between T and tf (about 6 centiMorgans proximal to tf) compared with its usual location about 10–15cM distal to tf . The simplest explanation for this finding would be a chromosomal inversion in the t -haplotypes. In the *SD* system of *Drosophila*, suppression of recombination is known to be associated with inversions. However, as yet, no karyologically detectable inversions associated with t -bearing chromosomes have been found. Thus the authors suggest a chromosomal transposition; unfortunately the haplotypes so far studied are deficient in polymorphic variants of other loci, the study of which could help to clarify this problem.

Despite the early progress made by Lyon and her colleagues the suppression of recombination associated with the t -haplotypes bedevilled attempts to make a detailed dissection of the pleiotropic effects of these mutant chromosomes into their component parts. The new finding that high-frequency recombination may be obtained between different t -haplotypes should now permit the isolation, in congenic stocks, of the various mutant loci of the t -haplotypes. These will become important tools in analysing how the different effects of the T/t complex on segregation, infertility and embryonic death

come about, and whether any of these have common genetic causes. For example, coupled with the development of monoclonal antibodies they should allow a clear analysis of whether any of the effects are mediated by specific cell-surface molecules.

The results also suggest new ways for investigating the genetic organization of the t -haplotypes. For example, the finding that the $H-2$ locus lies close to the lethal factors for t^{w5} , t^{w18} and t^{w32} means that cloning of these genes using cosmids and 'walking' from the $H-2$ region is a feasible proposition. □

In search of the Hubble parameter

from Sidney van den Bergh

NEW techniques of determining the distances of galaxies, and recent results using some of the more traditional methods of distance determination have, as de Vaucouleurs points out in this issue of *Nature* (p303), resulted in discordant estimates of the Hubble parameter. These apparent discrepancies are important because the global Hubble parameter is a measure of the scale-size and hence the age of the Universe.

Efforts to calibrate the extra-galactic distance scale have proceeded along two parallel tracks. One class of methods employs objects of known luminosity within galaxies, such as Cepheids, novae and the brightest red giants, as standard candles. A second more robust approach uses the galaxies themselves as luminosity indicators. Baum¹ first showed that the colours and luminosities of elliptical galaxies were correlated and subsequently van den Bergh^{2,3} devised a system of luminosity classification for spirals. Both methods suffer from the relatively low precision with which the luminosities of individual galaxies can be determined and further uncertainties in the extent to which environmental factors, such as membership in clusters, might affect the colours of ellipticals or the morphology of spirals. A particularly exciting new technique for the distance determination of spiral galaxies stems from the recent discovery by Tully and Fisher⁴ that the luminosity of a spiral is highly correlated

with its 21-cm line width. The correlation between velocity dispersion and luminosity of ellipticals also promises to provide distances for early-type galaxies.

The nature of the controversy that has arisen from these new measurements is perhaps best illustrated by comparing the results of three of the best-known recent determinations of the Hubble parameter.

De Vaucouleurs^{6,7} used traditional distance indicators such as Cepheids, RR Lyrae stars and novae to calibrate the 21-cm line width versus blue luminosity relation for galaxies. Aaronson, Mould and Huchra⁸⁻¹⁰ used the nearby Local Group galaxies M 31 and M 33, which have very well-determined distances, to calibrate the 21-cm line width versus infrared luminosity relation for spirals. Finally, Sandage and Tammann¹¹ used the luminosity of the brightest red supergiants (calibrated using Cepheids) to determine the maximum luminosity of type I supernovae. The apparent brightnesses of SN I then allowed them to find the distances to galaxies in and beyond the Virgo Cluster. The results obtained by these three techniques are compared in Table 1.

The distances for the relatively nearby M

Sidney van den Bergh is the Director of the Dominion Astrophysical Observatory, 5071 West Saanich Road, Victoria, BC V8X 4M6, Canada.

Table 1 Comparison of Hubble parameters and distances

	Aaronson, Mould & Huchra ⁸⁻¹⁰	De Vaucouleurs ^{6,7}	Sandage & Tammann ¹¹
Distance to:			
NGC2403/M81 group	3.6 ± 0.3 Mpc	3.2 Mpc	3.25 Mpc
M 101 group	6.8 ± 1.0 Mpc	5.0 Mpc	6.9 Mpc
Virgo Cluster	15.7 ± 0.6 Mpc	15.0 ± 0.6 Mpc	21.7 ± 3.1 Mpc
H_0 (local)	62 ± 4 km s ⁻¹ Mpc ⁻¹	64 km s ⁻¹ Mpc ⁻¹	45 ± 7 km s ⁻¹ Mpc ⁻¹
H_0 (global)	95 ± 4 km s ⁻¹ Mpc ⁻¹	95 km s ⁻¹ Mpc ⁻¹	50 ± 7 km s ⁻¹ Mpc ⁻¹
Infall velocity	509 ± 99 km s ⁻¹	250 km s ⁻¹	118 ± 223 km s ⁻¹

- Artzt, K., McCormick, P. & Bennet, D. *Cell* **28**, 463–470 (1982).
- Artzt, K., Shin, H.-S. & Bennett, D. *Cell* **28**, 471–476 (1982).
- Dobrovolskaia-Zawadska, N. *Cr. Séanc. Soc. Biol.* **97**, 114–119 (1972).
- Dobrovolskaia-Zawadska, N. & Kobozieff, N. *Cr. Séanc. Soc. Biol.* **110**, 728–784 (1932).
- Chesley, P. & Dunn, L. C. *Genetics* **21**, 525–526 (1936).
- Bennett, D. *Science* **144**, 263–267 (1964).
- Guenet, J.-L., Condamine, H., Gaillard, J. & Jacob, F. *Genet. Res.* **36**, 211–217 (1980).
- Gluecksohn-Waelsch, S. & Erickson, R. P. *Curr. Topics dev. Biol.* **5**, 281–316 (1970).
- Bennett, D. *Cell* **6**, 441–454 (1975).
- Sherman, M. I. & Wudl, L. R. in *Concepts in Mammalian Embryogenesis* (ed. Sherman, M. I.) 136–234 (MIT Press, 1977).
- Gluecksohn-Waelsch, S. & Erickson, R. P. *Proc. Symp. Immunogenet. of the H-2 System*, Liblice-Prague, 120–122 (1970).
- Yanagisawa, K., Bennett, D., Boyse, E. A., Dunn, L. C. & Dimeo, A. *Immunogenetics* **1**, 57–67 (1974).
- Artzt, K., Bennett, D. & Jacob, F. *Proc. natn. Acad. Sci. U.S.A.* **71**, 811–814 (1974).
- Marticorena, P., Artzt, K. & Bennett, D. *Immunogenetics* **7**, 337–347 (1978).
- Cheng, C. C. & Bennett, D. *Cell* **19**, 537–543 (1980).
- Goodfellow, P. N., Levinson, J. R., Gable, R. J. & McDevitt, H. O. *J. Reprod. Immun.* **1**, 11–21 (1979).
- Ginsberg, L. & Hillman, N. *J. Embryol. exp. Morph.* **33**, 715–723 (1975).
- Wudl, L. R. & Sherman, M. I. *Cell* **9**, 523–531 (1970).
- Lyon, M. F. & Phillips, R. J. G. *Heredity* **13**, 23–32 (1959).
- Lyon, M. F. & Meredith, R. *Heredity* **19**, 301–312 (1964).
- Lyon, M. F. & Meredith, R. *Heredity* **19**, 313–325 (1964).
- Lyon, M. F. & Meredith, R. *Heredity* **19**, 327–330 (1964).
- Lyon, M. F. & Mason, I. *Genet. Res.* **29**, 255–266 (1977).
- Styrna, J. & Klein, J. *Genet. Res.* **38**, 315–325 (1981).
- Sandler, L., Hiraizumi, Y. & Saudler, I. *Genetics* **44**, 232–250 (1959).
- Hartl, D. L. *Genetics* **73**, 613–629 (1973).
- Braden, A. W. H. et al. in *Edinb. Symp. Genetics of the Spermatozoon* (eds Beatty, R. A. & Gluecksohn-Waelsch, S.) 310–312 (1982).
- Silver, L. M. & Artzt, K. *Nature* **290**, 68–70 (1981).
- Lyon, M. F., Evans, E. P., Jarvis, S. E. & Sayers, I. *Nature* **279**, 38–42 (1979).
- Silver, L. M., White, M. & Artzt, K. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6077–6080 (1980).

81 and M 101 groups are seen to be in reasonable agreement with each other. Differences first appear at the distance to the Virgo Cluster where many kinds of distance indicators begin to fade below the plate limit. Errors quoted for each distance do not fully take into account external and systematic errors so it is probable the distances obtained for the Virgo Cluster are actually consistent with one other. Hopefully, observations of novae in Virgo ellipticals obtained with CCD cameras and of Cepheids with the space telescopes will soon reduce the remaining uncertainty in the distance to the Virgo Cluster.

From 160 radial velocities of Virgo Cluster members, Kraan-Kortweg¹² finds a mean velocity (corrected for solar motion) $V_0 = 967 \pm 53 \text{ km s}^{-1}$ from which the values of H_0 (local) given in the table may be derived. Many, but not all, available data appear consistent with H_0 (local) = $55 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

Because clusters of galaxies contain so much 'missing mass' they exert a significant gravitational pull on galaxies in their neighbourhood. As a result, our Milky Way System appears to be falling towards the Virgo Cluster. Direct evidence for even larger-scale anisotropies in the galaxian velocity field is provided by giant holes¹³ in the distribution of galaxies. Peebles¹⁴ points out that velocity anomalies at the edges of such holes probably lie in the range 100 to 350 km s^{-1} .

To minimize the effects of such deviations from a smooth Hubble flow the global value of the Hubble parameter has to be determined from distant galaxies that are situated well beyond the Local Supercluster. In such distant galaxies, standard candles are dim and standard meter sticks have characteristic sizes close to that of stellar seeing disks. Accurate measurement of the distances to such objects is therefore one of the most difficult problems in observational astronomy. Traditionally only the most daring and optimistic observational cosmologists have ventured into this difficult area. It is then surprising that some of them are, at times, also a bit optimistic concerning the influence of various selection effects and about their own observational errors? It is this that probably explains the apparent discrepancies in published values of H_0 (global). □

In vitro mutagenesis

from Tim Harris

CLONING techniques are now allowing the isolation and direct structural analysis of a large number of prokaryotic and eukaryotic genes and their regulatory regions, without recourse to classical genetic techniques. The ability to introduce changes into these cloned DNAs at predetermined sites and to assess the effect of these changes biologically (sometimes described as 'genetics in reverse') is rapidly becoming a routine part of gene manipulation technology. This was demonstrated quite clearly at a recent Cold Spring Harbor symposium entitled 'In Vitro Mutagenesis'.

In theory, the present techniques allow deletions, insertions or point mutations to be introduced at almost any location in a cloned DNA molecule¹. In practice, the type of mutation introduced depends largely on what is already known about the DNA and what biological questions are being asked. Several groups, for example, have analysed the promoter regions of cloned eukaryotic genes by the successive deletion of sequences upstream of the coding region, using exonuclease III and nuclease S_1 or the double strand - specific nuclease *Bal*-31. The effect of the deletion on the level of transcription of the gene can then be measured either *in vitro* or *in vivo*.

One disadvantage of exonucleases commonly used to generate deletions is that they act at both ends of a linear DNA molecule simultaneously. Unidirectionality can be achieved, however, by incorporating an α -thio deoxynucleotide at the 3' end of one strand to prevent the action of exonuclease III (ref. 2). The functional domains of the alanyl-tRNA synthetase gene are being examined in this way by deletion of 3' portions of the coding sequence without affecting the 5' end of the gene³.

In addition to deletion mutations, single base changes can also be generated at or near restriction enzyme sites. In the presence of ethidium bromide, some restriction enzymes nick DNA in one strand rather than cutting both strands. This nick can be extended to a gap by exonuclease treatment and base transitions introduced into the remaining single-stranded DNA by bisulphite treatment⁴. Bisulphite specifically deaminates cytosine to uracil in single stranded nucleic acids⁵, resulting in a transition mutation during repair. Obviously, restriction sites are not always conveniently placed for gap or deletion mutagenesis. This limitation has been overcome to some extent by utilizing the fact that in the presence of *Escherichia coli* RecA protein, small single-stranded DNAs will efficiently displace a complementary sequence in a supercoiled double-stranded DNA molecule forming a D-loop. The displaced single-stranded

loop is then available for mutagenesis using nuclease S_1 and bisulphite⁶. In this way a number of point mutations can be introduced into a target area defined by the size of the DNA used to form the D-loop. The ease of cloning DNA in single-stranded phage cloning vectors (such as M13) has extended the versatility of this approach; the formation of specific heteroduplexes allows the displacement and mutagenesis of defined regions of DNA and the efficient recovery of the mutants by transfection. Mutations have been introduced into the early promoters of both SV40⁷ and adenovirus⁸ by this technique and this kind of segment-specific mutagenesis has also been used to analyse functional sequences in the *lac* promoter⁹.

Besides introducing mutations into single-stranded DNA directly, mutations can also be introduced by misincorporation of nucleotides during repair synthesis with DNA polymerase. The simplest system is to nick (and gap) at a restriction site and 'fill in' with DNA polymerase in the absence of a nucleotide or in the presence of a nucleotide analogue¹⁰. Base exchange during repair can be prevented either by incorporation of α -thio deoxynucleotides¹¹ or by using reverse transcriptase, which lacks a proofreading activity. DNA ligase is used to freeze mutations after repair by sealing across the nick.

There is no doubt that the most powerful method for obtaining defined single base changes is the oligonucleotide-directed method championed by M. Smith and applied initially to the single-stranded phage ϕ X174¹². In this method, an oligonucleotide with a mismatch or small deletion is annealed to a single-stranded template and used to prime the synthesis of a complementary strand. Theoretically, 50 per cent of the progeny molecules arising after transfection should be mutant. The power of the method lies in the fact that there is virtually no constraint on the target, provided the nucleotide sequence of the DNA is known. Moreover, the identification of the progeny is now relatively straightforward since the primer which was used to introduce the mutation can also be used as a probe to search for mutants under stringent hybridization conditions¹³. The method is also very versatile; it can be used on plasmid DNA rendered single stranded by nicking and extensive exonuclease III digestion as well as on DNA cloned in M13 or other single-stranded vector. One possible drawback is the need for a supply of well characterized synthetic oligonucleotides. If nothing else,

Tim Harris is a group leader in Molecular Biology at Celltech Ltd, Slough, UK.

1. Baum, W.A. *Publ. astr. Soc. Pacif.* **71**, 106 (1959).
2. van den Bergh, S. *Astrophys. J.* **131**, 215 (1960a).
3. van den Bergh, S. *Astrophys. J.* **131**, 558 (1960b).
4. Tully, R.B. & Fisher, J.R. *Astr. Astrophys.* **54**, 661 (1977).
5. Faber, S.M. & Jackson, R.E. *Astrophys. J.* **204**, 668 (1976).
6. De Vaucouleurs, G. *Astrophys. J.* **224**, 710 (1978).
7. De Vaucouleurs, G. *Astrophys. J.* **253**, 520 (1982).
8. Aaronson, M., Mould, J. & Huchra, J. *Astrophys. J.* **237**, 655 (1980).
9. Mould, J., Aaronson, M. & Huchra, J. *Astrophys. J.* **238**, 458 (1980).
10. Aaronson, M. *et al.* *Astrophys. J.* **239**, 12 (1980).
11. Sandage, A. & Tammann, G.A. *Astrophys. J.* **256**, 339 (1982).
12. Kraan-Kortweg, R.C. *Astr. Astrophys.* **2**, 04, 280 (1981).
13. Kirshner, R.P. *et al.* *Astrophys. J. Lett.* **248**, L57 (1981).
14. Peebles, P.J.E. *Astrophys. J.* **257**, 438 (1982).

this should provide an additional stimulus to the development of rapid oligonucleotide synthesis methods.

Several examples of the application of oligonucleotide site-directed mutagenesis are available. It has been used to resolve the functions of two overlapping spliced adenovirus mRNAs made early in infection — a question which could not have been answered unequivocally by random or target-directed mutagenesis¹⁴. At least two groups have used the procedure to alter the anticodon of a cloned tRNA gene (for example, tRNA^{Lys} or tRNA^{Tyr}) to that of a suppressor tRNA gene. The function of the tRNA was then tested by its ability to suppress various amber mutations (see, for example, ref. 15). Finally, the method has been used to probe the function of the signal peptide present on secreted proteins. Changing the charge of the amino acids at the NH₂ terminus of the *E. coli* outer membrane protein affected the ability of the protein to be assembled in the membrane¹⁶.

Clearly, there are a wide variety of biological problems which will benefit from the application of both target- and site directed *in vitro* mutagenesis. It is now

possible to introduce specific changes into cloned DNA; when coupled with high-level gene expression vectors it should also be possible to assay the effect of these changes in several prokaryotic and eukaryotic systems. For certain proteins, for which there is detailed structural information and a cloned gene available, it is not too fanciful to consider that specific mutations could be introduced to produce molecules with altered or novel activities. □

1. Shortle, D., DiMaio, D. & Nathans, D. *A. Rev. Genet.* **15**, 265-294 (1981).
2. Putney, S.D., Benkovic, S.J. & Schimmel, P.R. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7350-7304 (1981).
3. Jasin, M. & Schimmel, P. Abstract of the meeting.
4. Shortle, D. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2170-2174 (1978).
5. Hayatsu, H. *Prog. Nucleic Acid Res. molec. Biol.* **16**, 75-124 (1976).
6. Shortle, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 5375-5379 (1980).
7. Everett, R.D. & Chambon, P. *EMBO J.* **1**, 433-437 (1982).
8. Hearing, P. & Schenk, T. Abstract of the meeting.
9. Weiher, H. & Schaller, H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1408-1412 (1982).
10. Müller, W. *et al. J. molec. Biol.* **124**, 343-358 (1978).
11. Shortle, D. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 1588-1592 (1982).
12. Smith, M. & Gillam, S. *Genet. Engng* **3**, 1-32 (1981).
13. Wallace, R.B. *et al. Nucleic Acids Res.* **9**, 3647-3656 (1981).
14. Montell, C. *et al. Nature* **295**, 380-384 (1982).
15. Temple, G.F. *et al. Nature* **296**, 537-540 (1982).
16. Inouye, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **79**, 34438-3441 (1982).

Crown gall tumours and plant growth regulators — any connection?

from Mike Bevan

CROWN gall tumour cells produced by infection with *Agrobacterium tumefaciens* can grow in culture in the absence of endogenous auxins such as indole acetic acid (IAA), in contrast to normal plant cells. Transformation takes place when part of the giant Ti, or tumour-inducing, plasmid is stably integrated into the nuclear DNA of the host plant. It is not clear whether the tumorous state is caused by the insertion of T-DNA into plant sequences, or if T-DNA encodes control or structural genes for plant growth regulators. Recently, Lui *et al.*¹ presented evidence for the presence of an IAA gene on the Ti plasmid. Although mutations in the IAA locus render the bacteria harbouring the plasmid avirulent, the locus itself is in a part of the plasmid that is not stably maintained in tumours.

The gene, *iaaP*, responsible for IAA synthesis was identified on the tumour-inducing plasmid by inserting the plasmid into an avirulent, plasmid-free mutant that did not synthesize IAA (*A. tumefaciens* cells themselves normally produce large amounts of IAA). The *iaaP* locus, which may encode enzyme(s) responsible for the conversion of indolepyruvate to IAA, was

mapped by transposon (Tn5) mutagenesis to a region of the Ti plasmid known to be involved in oncogenesis. These sequences, called the virulent region, have been delineated by transposon mutagenesis to a region of about 30 kilobases (in pTiB6806^{2,3}) which may mean that many traits other than IAA production are involved in virulence. It remains to be demonstrated that this Tn5 insertion does not have polar effects on adjacent virulence genes. The question now remains, how do these virulence genes, in particular the *iaaP* locus, incite tumour formation when it is only the T-DNA, some 20 kb from the virulent region, that is stably maintained in the plant tumour?

Lui *et al.* have three suggestions. First, IAA secreted by virulent bacteria may 'precondition' a wounded plant cell by inciting cell wall changes and DNA synthesis which would allow subsequent transformation. This is reminiscent of gall induction on olive and oleander stems by *Pseudomonas savastanoi*, which harbours a plasmid-encoded enzyme that converts indoleacetic acid to IAA. Virulence, as measured by gall formation, is directly related to IAA levels⁴, and no transfer of plasmid DNA to the host plant has been observed. Second, the authors speculate that IAA may be a 'second messenger' that activates other virulence genes when the bacterium

encounters a suitable plant cell. Finally, it is suggested that some sort of DNA arrangement may integrate the IAA gene(s) into T-DNA in an early transient stage of crown gall formation.

Recent evidence from extensive transposon mutations in T-DNA indicates that several genes in a highly conserved region of T-DNA are involved in tumour growth and morphology. Two genes, when mutated, give rise to a shooty phenotype, and one gene gives rise to a rooty phenotype when mutated⁵. These findings can be formally explained by postulating that the rooty locus regulates cytokinin levels, and the two shooty loci regulate auxin (IAA?) levels. Ooms *et al.*⁶ demonstrated that exogenous auxin reversed the shooty phenotype and exogenous cytokinin reversed the rooty phenotype in mutated octopine tumours. Interestingly, nearly all of T-DNA can be deleted without affecting the ability of remnant left and right border fragments to be transferred into plant DNA⁷. These 'transformed' cells are, of course, non-tumorous. The virulence region, which maps outside T-DNA, has been implicated in the transfer of T-DNA to plant cells.

In the light of these findings there may be no need to postulate the integration of the non-T-DNA auxin gene into plant DNA to explain oncogenesis, as Lui *et al.* suggest. Indeed, the significance of IAA production by *Agrobacteria* is unclear as in the absence of endogenous tryptophan, both virulent and avirulent mutants failed to accumulate IAA⁸. Perhaps there is adequate tryptophan present in the cell sap of the wound where gall formation is initiated to induce IAA synthesis by virulent *Agrobacteria*, but this has not been investigated.

What does seem clear is that it is profitable to consider auxin and cytokinin autotrophy, regulated by loci on conserved sequences in T-DNA, as being important, if not the basis, for the autonomous growth of crown gall tumours. Endogenous growth regulator levels in plant cells are known to vary widely, and it is reasonable to assume that both sources of growth regulators contribute to the final growth rate and morphology of tumours. The search for a definite connection between the shooty and rooty loci and growth regulator production has been confounded by the very low concentrations of mRNA and proteins encoded by this region of T-DNA. A promising approach has been to direct the expression of T-DNA-encoded genes in *Escherichia coli* minicells⁹, and identify the activity of T-DNA-encoded proteins. □

1. Lui, S.-T., Perry, K.L., Shardt, L.L. & Kado, C.I. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2812-2816 (1982).
2. Garfinkel, D.J. & Nester, E.W. *J. Bact.* **144**, 732-743 (1980).
3. Ooms, G., Klapwijk, P.M., Poulsen, J.A. & Schilperoort, R.A. *J. Bact.* **144**, 82-91 (1980).
4. Smidt, M.L. & Kosuge, T. *Physiol. Pl. Path.* **13**, 203-214 (1978).
5. Garfinkel, D.J. *et al. Cell* **27**, 143-153 (1981).
6. Ooms, G., Hooykaas, P.J.J., Moolenaar, G. & Schilperoort, R.A. *Gene* **14**, 33-50 (1981).
7. Leemans, J. *et al. EMBO J.* **1**, 147-152 (1982).

Mike Bevan is in the Plant Breeding Institute, Mavis Lane, Trumpington, Cambridge CB2 2LQ.

The American kestrel as a laboratory research animal

from David M. Bird

AMERICAN KESTRELS (*Falco sparverius*) were first bred in the laboratory unintentionally by Morris¹ in 1850 and later gained acclaim as laboratory animals when they were used to help indict DDT as a significant cause of reproductive failure in wild birds². Since then, the kestrel has been used extensively in laboratory studies, for example, toxicology³, behaviour⁴ and physiology⁵, and has also proved useful as a model species to develop management techniques for endangered species such as the peregrine falcon (*Falco peregrinus*)⁶.

Two large captive colonies have been

established in North America, one in Laurel, Maryland for toxicological research⁷, and a pedigree one of over 200 individuals at McGill University, Montreal in a diverse research programme. Here, I shall outline some of the advantages of the American kestrel as a laboratory animal.

This kestrel nests commonly in most

David M. Bird is the Director of the Macdonald Raptor Research Centre of McGill University, 21,111 Lakeshore Rd, Ste Anne de Bellevue, Quebec H9X 1C0.

rural and urban areas. In North America, two subspecies exist: *F.s. sparverius* and *F.s. paulus*. *F. s. sparverius* weighs 85–140 g, the female being approximately one-third heavier than the male. This species can be easily sexed by plumage after 12 days of age: males have blue-grey wings and female rufous-brown. The kestrel is sexually mature in its first spring and breeds readily in captivity. Unmated captive-raised females can often be induced to lay eggs simply by providing a nestbox during the breeding season, thus facilitating studies on artificial insemination⁶. Ninety per cent of randomly selected pairs lay an egg every two to three days to produce a first clutch of five eggs. If the first clutch is removed within a week of the laying of the last egg, 95 per cent of the females recycle 11 days later to lay a second clutch⁸. Replacement clutches, though, may have one less egg, longer eggs and eggs with thicker shells than first clutches⁸. If eggs are removed as they are laid, up to 26 eggs can be produced.

The overall fertility of randomly selected pairs is about 65 per cent, but rises to 80 per cent in compatible pairs⁹. The kestrel can be easily inseminated artificially, as males massaged every two days will give, on average, 0.01 ml of semen containing about 400,000 spermatozoa⁶. The mean duration of fertility in females after one insemination averages 8 days with a maximum of 12 days¹⁰. Motile sperm can be obtained after freezing and thawing in a glycerol medium.

The breeding season usually begins in early April in temperate regions, peaks at about 13.5 h of daily light and is strongly influenced by the photoperiod. Pairs can be induced by changing the photoperiod to undergo an extra breeding season during freezing temperatures between two successful spring seasons¹¹. Hatchability with either natural or artificial incubation averages between 60 and 70 per cent, decreasing as the season progresses. The large variation in hatchability from year to year is being investigated¹². Eggs take 28–30 days to hatch. For artificial incubation, a temperature of 37.5°C, 50–60 per cent humidity and turning at least four times daily is ideal. Young hatching from artificially incubated eggs can be hand-reared or fostered to natural parents with eggs or young. Kestrels hand-reared in family groups do not exhibit abnormal reproductive behaviour when sexually mature, but hand-reared females do lay larger clutches with heavier eggs¹³. Hand-rearing, that is, feeding to satiation four times daily, should be used with caution, however, as it may result in physically smaller birds¹³. The sex ratio at hatching is 1:1. Nestlings take 25–30 days to fledge.

The McGill colony is sustained on an *ad libitum* diet of day-old cockerels, usually one per bird daily, supplemented by SA-37 (Rogar-STB Ltd, Montreal) vitamin/mineral supplements and dietary lime-



100 years ago

PROFESSOR HAECKEL IN CEYLON

THE great black scorpion (nearly a foot long) is so common in Ceylon that I once collected half a dozen in the course of an hour. Snakes exist also in great numbers. Slender green tree-snakes hang from almost every bough, and at night the great rat-snake (*Coryphodon Blumenbachii*) hunts rats and mice over the roofs of the huts. Although they are harmless and their bite not poisonous, it is by no means a pleasant surprise when one of these rat-snakes, five feet long, suddenly drops through a hole in the roof into one's room, occasionally alighting on the bed.

On the whole, however, my nights in Belligam were but little disturbed by animal intruders, although I was often kept awake by the howling of jackals and the uncanny cry of the Devil-bird (a kind of owl, *Syrnium Indrani*) and other night-birds. The bell-like cry of the pretty little tree-frogs which make their dwelling in the cups of large flowers, acted rather as a slumber song. But I was far oftener kept awake by the whirl of my own thoughts, by the recollection of the many events of the past day, and the anticipation of that which was to come. A brilliant succession of lovely scenes, of interesting observations and varied experiences mingled in my brain with plans of fresh enterprise and new discoveries for the morrow.

At seven o'clock my boatmen appeared to carry down my nets and glasses for the daily canoe expedition. This lasted from two to three hours, and on my return I busied myself in disposing my captures in glasses of different sizes, and saving such as could be saved among the few survivors. The more important specimens were microscopied and drawn at once. Then I had my second bath, and at eleven o'clock appeared my so-called 'breakfast', consisting chiefly of curry and rice. The rice was simply boiled, but in the preparation of the curry my old cook, Babua, exerted all the ingenuity with which nature had endowed his diminutive brain to present me with a fresh

combination every day. Sometimes the curry was 'sweet', sometimes 'hot'; sometimes it appeared as an undefinable *mixtum compositum* of vegetables, sometimes as a preparation of the flesh of various animals. Babua seemed to divine that as a zoologist I was interested in every class of animal life, and that he could not do better than turn my curry into a sort of daily zoological problem. . . . He was apparently a staunch upholder of the theory of the near relationship of birds and reptiles, and held it to be immaterial what particular species of *Saurian* were prepared for the table.

It is estimated by Prof. Dufour (*Arch. des Sciences*) that in a disastrous hailstorm on August 21 last year, about 100,000 cubic metres of ice fell in the district of Morges alone in a few minutes, and probably more than 1,000,000 cubic metres in the whole canton de Vaud that afternoon. Yet this is a small matter compared with the terrible hailstorm of July 13, 1788 (regarding which he makes some calculations). He gives some interesting facts, which seem to have been overlooked, in the history of *paragrêles*, or hail-preventers. Old men in the Canton de Vaud remember such apparatus, of lightning-rod character, being set up in several vineyards in 1825; the object being to hinder the formation of hail, by withdrawing electricity from the clouds. A hailstorm in July 1826 devastated, it is said, the best protected vineyards, and the *paragrêles* were then removed. Yet it was on receipt of encouraging and credible testimony from Italy and France (Prof. Dufour shows by extracts) that this brief experiment was made. Considering the distance of hail-forming clouds from the highest *paragrêles*, it is difficult, the author considers, to admit an influence of such apparatus; yet it must be remembered that electricity is "*un véritable fluide à surprises*"; often showing new and unexpected properties. Lately it is said to have been observed in some Swiss cantons, that showers of hail are more rare near forests than in unwooded districts. Prof. Dufour notes this as a matter calling for investigation. A forest may be regarded as a collection of *paragrêles*, and should it be proved to have the influence referred to, the theories which prevailed in 1824 and 1825 would gain new support.

From *Nature* 26, 503; 21 September 1882.

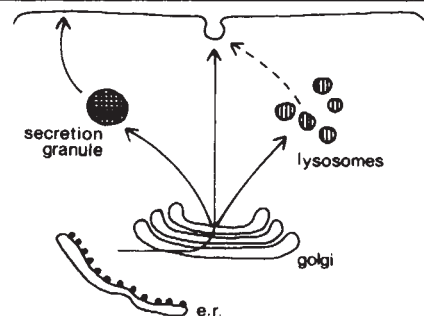
stone, and occasionally laboratory mice. Falcons do not require water for drinking, but do enjoy bathing. Pen design has varied from cages of wire mesh⁷, 15.2 × 6.1 × 1.8 m (L × W × H), to cages of polyethylene or cardboard, 1.5 × 1.2 × 1.2 m. The standard design now preferred at McGill is a plywood box, 2.4 × 1.2 × 2.4 m, with wire mesh roof and floor. Each pen is equipped with a nestbox, 25 × 25 × 46 cm, accessible by a trap-door, several sisal rope perches 1.3 cm in diameter and a wooden feeding perch 5 cm wide.

The colony is usually wintered in flocks of about 20 birds housed in unheated flight pens, 6.1 × 6.1 × 2.4 m, with rope perches. Annual mortality may be up to 10 per cent, and results mainly from intraspecific aggression and winter temperatures of -30°C. Several birds are still breeding successfully after 9 years in captivity.

Approximately 1 ml (10 per cent of total blood volume) of blood can be removed weekly from the brachial vein with no ill effects. Haematological data have been reported^{14,15} and assay techniques for determining the levels of six reproductive steroid hormones are almost complete in our laboratory. We are now investigating factors influencing mate selection in

falconiformes using mate choice tests¹⁶, as well as the influence of diet quality versus frequency of feedings during nestling growth on subsequent body size and reproductive performance. A pedigree colony enhances its value in genetic research for wild-type birds. A wild colony of at least 25 pairs of kestrels has been established using nestboxes within a 10 mile radius of our laboratory to permit comparative studies between wild and captive birds. The use of the McGill University colony is extended to other interested scientific institutions. □

1. Morris, F.O. *Zoologist*, 2648 (1850).
2. Wiemeyer, S.N. & Porter, R.D. *Nature* **227**, 737 (1970).
3. Lincer, J.L. & Peakall, D.B. *Nature* **228**, 783 (1970).
4. Mueller, H.C. *Nature* **233**, 345 (1971).
5. Fox, R. *et al. Science* **192**, 263 (1976).
6. Bird, D.M. *et al. Can. J. Zool.* **54**, 1183 (1976).
7. Porter, R.D. & Wiemeyer, S.N. *J. Wildl. Mgmt* **34**, 594 (1970).
8. Bird, D.M. & Lagüe, P.C. *Can. J. Zool.* **60**, 71 (1982).
9. Bird, D.M. & Lagüe, P.C. *Can. J. Zool.* **60**, 80 (1982).
10. Bird, D.M. & Buckland, R.B. *Can. J. Zool.* **54**, 1595 (1976).
11. Bird, D.M. *et al. Can. J. Zool.* **58**, 1022 (1980).
12. Bird, D.M. in *Recent Advances in Raptor Diseases* (Chiron, UK, 1981).
13. Bird, D.M. & Lagüe, P.C. *Can. J. Zool.* **60**, 89 (1982).
14. Snyder, J. *et al. in Recent Advances in Raptor Diseases* (Chiron, UK, 1981).
15. Rehder, N.B. *et al. Comp. Biochem. Physiol.* **72A**, 105 (1982).
16. Bird, D.M. & Goldblatt, C. in *Recent Advances in Raptor Diseases* (Chiron, UK, 1981).



In eukaryotic cells, integral plasma membrane proteins and proteins destined for secretion share a common origin: translation on membrane-bound polysomes concomitant with vectorial insertion into (or across) the membrane of the endoplasmic reticulum (ER)^{18,21}. The first post-translational modification of these proteins, glycosylation, also occurs at this stage²². This often involves the transfer of a branched, mannose-rich oligosaccharide to an asparagine residue on the nascent peptide. Processing of the carbohydrate chain begins within minutes of its addition. First, glucose residues are removed, followed sequentially by the trimming of several mannoses and the addition of a new complement of terminal sugars (GlcNAc, fucose, galactose and sialic acid) via the activity of several different glycosyl transferases. Terminal glycosylation and at least some mannose removal occurs following the transfer of newly synthesized proteins from the ER to the Golgi². Membrane vesicles then form which transport lysosomal, secretory and plasma membrane proteins to their appropriate destinations.

Multiple pathways of membrane transport

from Ira Mellman

INSIDE the cell a complex pattern of membrane traffic is generated by the intake and secretion of proteins and the manufacture and recycling of membrane components. Recent experiments suggest that the biosynthesis and transport of membrane, lysosomal and certain secretory proteins have a great deal in common, and may share parts of the same constitutive pathway. At the same time, there is also biochemical evidence that at least those secretory proteins which are sequestered in secretion granules before exocytosis are transported along a distinct pathway once they have left the Golgi apparatus.

The Golgi apparatus plays a central part in the post-translational processing of the carbohydrate moieties of membrane and secretory glycoproteins. It is a complex structure consisting of several flattened cisternae, usually with dilated rims, and its mode of operation is the topic of considerable debate. Central to this question is an understanding of the pathway(s) of membrane transport into, through and out from the Golgi stack. Traditionally, movement through the Golgi has been thought to be in one direction, with newly synthesized proteins (both membrane and secretory) entering at the *cis* face (the side defined

morphologically as being closest to the endoplasmic reticulum (ER) and/or nucleus) and exiting as mature glycoproteins at the opposite or *trans* face¹. While the existence of such directionality is not yet proved, Rothman² has recently postulated that individual cisternae of the stack could act like units of a distillation apparatus. Proteins destined for transport to the plasma membrane would be successively processed and enriched relative to contaminants (for example, components of the ER membrane) as the former move by vesicular transport from *cis* to *trans* cisternal elements.

Recent experiments from Rothman's laboratory³ suggest the existence of at least two temporally and functionally distinct components within the Golgi which are involved in the processing of carbohydrates of membrane glycoproteins. Homogenates of CHO cells were centrifuged in sucrose density gradients and, among the fractions, the putative 'early-acting' processing enzyme α_2 -mannosidase, which trims mannose residues from the oligosaccharide, was

found to be slightly separated from the 'late-acting' enzymes involved in terminal glycosylation, galactosyl transferase and sialyl transferase (see Fig. 1 legend). Interestingly, when labelled homogenates of cells infected with vesicular stomatitis virus were centrifuged, the mature (that is, galactose-labelled) form of the viral spike glycoprotein G was found to co-sediment with galactosyl transferase, while a more immature form of G protein sedimented with mannosidase.

If, as these findings imply, early events, such as mannose-trimming, take place in one compartment while later stages, such as terminal glycosylation, take place in another, then it is appealing to speculate that the two compartments correspond to the *cis* and *trans* Golgi cisternae. Such an interpretation must, however, remain speculative given the lack of available corroborative morphological data. Although it has been demonstrated that cisternal elements are cytochemically heterogeneous, most of the activities relevant to glycosylation have yet to be localized to *cis* or *trans* Golgi elements. Indeed, evidence for heterogeneity within single cisternae has been found^{4,5}. In addition, it is also not at all certain to what extent Golgi subcompartments retain their integrity after homogenization.

Further progress requires specific markers that will permit the separation and

Ira Mellman is Assistant Professor of Cell Biology in the School of Medicine, Yale University, New Haven, Connecticut 06510.

identification of Golgi subcompartments, particularly in experiments in cell-free systems⁶⁻⁸ where the morphological features which characterize the Golgi are usually lost.

Some first steps in this direction come from recent experiments by Roth and Berger in which a terminal glycosyltransferase has been immunocytochemically localized to the *trans* aspect of the Golgi apparatus of HeLa cells⁹. An antiserum against human milk galactosyltransferase was used and the antigen visualized on fixed, embedded material by protein A-colloidal gold staining. Gold was localized to two or three cisternae, coincident with a second Golgi cytochemical marker thiamine pyrophosphatase that is usually detected in *trans* elements. However, the '*trans-most*' Golgi cisternae, marked cytochemically by acid phosphatase activity, were not labelled with gold.

A novel approach to the identification of Golgi subcompartments has recently been taken by Louvard *et al.*¹⁰. They immunized rabbits with a rat liver Golgi fraction and obtained an antiserum which exhibited significant anti-Golgi activity. Following sequential adsorptions against rat blood plasma and microsomal membranes, the serum showed highly specific Golgi reactivity as judged by immunofluorescence and by immunoperoxidase labelling in the electron microscope. The antiserum labelled many, but not all, of the stacks of the Golgi complex in rat NRK cells. Whether these were *cis* or *trans* cisternae could not be determined. Many Golgi-associated vesicles were also labelled, some of which had clathrin coats. The antigenic determinant appeared to be on the luminal surface of both cisternae and vesicles. Rather surprisingly, the antiserum recognizes only a single 135,000 molecular weight polypeptide, in both NRK cells and rat liver. While this particular protein may not prove to be useful for Golgi isolation and identification in cell fractions (for example, by immunoabsorption, as demonstrated by Ito and Palade¹¹), the value of the approach is clear. In fact, another paper¹² in the same issue of *Journal of Cell Biology* reports the isolation of a monoclonal antibody against another Golgi-associated protein. A crude (Triton-insoluble) extract of cells was used as the immunogen. However, the antibody's specificity was defined only at the level of immunofluorescence, thus the actual localization of its antigen must await further study.

The fact that both membrane and secretory proteins undergo proximal and terminal glycosylation implies that they may take the same or similar routes through the Golgi apparatus. Once proteins have left the Golgi, though, it seems membrane and secretory proteins do not always take the same pathway to the plasma membrane (Fig. 1). Two general types of secretory activity can be dis-

tinguished. The first, which presumably occurs in many cell types, mediates an apparently continuous and non-concentrative exocytosis of secretory products. In the second, newly synthesized secretory proteins are concentrated and stored in intracellular granules, as in exocrine cells. The granules can be induced to fuse with the plasma membrane by specific secretagogue stimuli. This latter pathway is unlikely to be responsible for delivering new plasma membrane components to the cell surface since the polypeptide composition of the secretion granule is considered quite 'simple' and turns over more slowly than the contents it delivers¹³.

Direct evidence consistent with the existence of at least two separate pathways has now been provided by Gumbiner and Kelly¹⁴. They have studied a pituitary tumour cell line (AtT-20) which synthesizes the polypeptide hormone ACTH and the gp70 envelope glycoprotein of an endogenous murine leukaemia virus. Both are glycosylated during synthesis. Proteolytically processed (mature) ACTH is sequestered intracellularly in secretion granules¹⁵ and is released only slowly in the absence of secretagogues ($t_{1/2} = 3-4$ h). In contrast, gp70 appears on the cell surface and/or is released (secreted?) much more rapidly ($t_{1/2} < 30$ min). Incompletely processed ACTH precursors (polypeptides of molecular weights 23,000 and 30,000) were also continuously released with kinetics identical to gp70. These presumably represent hormone which escaped complete proteolytic processing and condensation in granules.

The differential rates of gp70 and mature ACTH release suggest that these molecules are transported by distinct mechanisms. Supporting evidence comes from the effect of secretagogues such as cyclic AMP, which cause a rapid increase in the rate of

ACTH secretion but have relatively little effect on ACTH precursor or gp70 release. Similarly, the membrane of isolated secretion granules was found to contain only an insignificant fraction of total cellular gp70. It obviously must be shown that the same cells actually secrete both ACTH and gp70, but these observations do suggest the existence of at least two separate pathways from the Golgi to the cell surface. Mature ACTH is probably transported by a route found only in specialized secretory cells. In contrast, gp70 transport may occur via a pathway which is involved in the continuous delivery of secretory products and new (or recycling) membrane components to the cell surface.

Recent evidence has also suggested that, at least in certain conditions, newly synthesized lysosomal hydrolases take the continuous route followed by gp70 and ACTH precursors. Like membrane and secretory proteins, lysosomal hydrolases are synthesized and glycosylated on the ER and then processed during transit through the Golgi complex^{16,17}. However, as shown by Kornfeld, Von Figura and others, processing results in the attachment of terminal mannose phosphate residues¹⁸. These are recognized by a specific membrane receptor which may be involved in the delivery of mannose phosphate-containing enzymes to lysosomes¹⁹. The delivery system seems only partially efficient since some newly synthesized hydrolases are usually secreted into the medium. Also, as Rosenfeld *et al.*¹⁷ show, virtually all the cathepsin D and β -glucuronidase synthesized by cultured rat hepatocytes are secreted when glycosylation is inhibited by tunicamycin.

Thus, lysosomal hydrolases may be transported from the Golgi via two distinct pathways, the first directed towards lysosomes, the other towards the cell surface (Fig. 1). Tunicamycin might inhibit the former route by preventing the formation of the mannose phosphate recognition marker. An alternative explanation, however, might be that most vesicles coming from Golgi which are bound for the plasma membrane first fuse with lysosomes (or another acidified vacuole) before continuing to the cell surface. Since low pH is known to dissociate mannose phosphate glycoproteins from their receptor, this pathway could still account for the delivery of newly synthesized hydrolases to lysosomes. Moreover, this interpretation is appealing since it provides an interface between the pathways of membrane transport during lysosomal biogenesis and during endocytosis, the latter being a process which generates an extensive bidirectional flow of membrane between lysosomes and the cell surface. Endocytosis is already known to transport a wide variety of plasma membrane proteins²⁰, including cell surface mannose phosphate receptors, which mediate the uptake and delivery of extracellular hydrolases to lysosomes¹⁹. □

1. Farquhar, M.G. & Palade, G.E. *J. Cell Biol.* **99**, 77s (1981).
2. Rothman, J.E. *Science* **213**, 1212 (1981).
3. Dunphy, W.G., Fries, E., Urbani, L.J. & Rothman, J.E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7453 (1981).
4. Farquhar, M.G., Bergeron, J.J.M. & Palade, G.E. *J. Cell Biol.* **60**, 8 (1974).
5. Cheng, H. & Farquhar, M.G. *J. Cell Biol.* **70**, 660 (1976).
6. Quinn, P.S. & Judah, J.D. *Biochem. J.* **172**, 301 (1978).
7. Bergeron, J.J.M., Rachubinski, R.A., Sikstrom, R.A., Posner, B.J. & Paiement, J. *J. Cell Biol.* **92**, 139 (1982).
8. Rothman, J.E. & Fries, E. *J. Cell Biol.* **89**, 162 (1981).
9. Roth, J. & Berger, E.G. *J. Cell Biol.* **93**, 223 (1982).
10. Louvard, D., Reggio, H. & Warren, G. *J. Cell Biol.* **92**, 92 (1982).
11. Ito, A. & Palade, G.E. *J. Cell Biol.* **79**, 590 (1978).
12. Lin, J.J.-C. & Queally, S.A. *J. Cell Biol.* **92**, 108 (1982).
13. Meldolesi, J., Borgese, N., DeCamilli, P. & Ceccarelli, B. in *Membrane Fusion* (eds Poste, G. & Nicolson, G.L.) 509 (Elsevier, Amsterdam, 1978).
14. Gumbiner, B. & Kelly, R.B. *Cell* **28**, 51 (1982).
15. Gumbiner, B. & Kelly, R.B. *Proc. natn. Acad. Sci. U.S.A.* **78**, 318 (1981).
16. Erickson, A.H. & Blobel, G. *J. biol. Chem.* **254**, 11771 (1979).
17. Rosenfeld, M.G., Kreibich, G., Popov, D., Kato, K. & Sabatini, D.D. *J. Cell Biol.* **93**, 135 (1982).
18. Sabatini, D.D., Kreibich, G., Morimoto, T. & Adesnik, M. *J. Cell Biol.* **92**, 1 (1982).
19. Gonzalez-Noriega, J.H., Grubb, Talkad, V. & Sly, W.S. *J. Cell Biol.* **85**, 839 (1980).
20. Mellman, I.S., Steinman, R.M., Unkeless, J.C. & Cohn, Z.A. *J. Cell Biol.* **87**, 712 (1980).
21. Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496 (1980).
22. Hanover, J.A. & Lennarz, W.J. *Archs Biochem. Biophys.* **211**, 1 (1981).

ARTICLES

Five crucial tests of the cosmic distance scale using the Galaxy as fundamental standard

G. de Vaucouleurs

Department of Astronomy and McDonald Observatory, University of Texas, Austin, Texas 78712, USA
Mount Stromlo and Siding Spring Observatories, Australian National University, Canberra

Five crucial tests of the 'long' and 'short' extragalactic distance scales based on a new approach using our Galaxy as the fundamental calibrator are presented. All confirm the short scale within 0.1–0.2 mag and the corresponding value of the Hubble constant $H_0 = 95 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$. All reject the long scale ($H_0 = 50 \text{ km s}^{-1} \text{ Mpc}$) at very high significance levels.

THE traditional approach to the extragalactic distance scale rests on a pyramid of primary, secondary and tertiary indicators of increasing range and decreasing accuracy. This multi-step procedure, fraught with the danger of cumulative errors, has led to two main, widely diverging scales: the 'long' scale¹ implying a Hubble constant $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$, and the 'short' scale² leading to $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$. Several authors have shown that the long scale rests on very precarious foundations²; counter-arguments have been offered in its defence and to criticize the short scale¹. A conclusive test was needed.

Such a test (actually five different ones) is now possible through direct quantitative comparison of our Galaxy with other galaxies: (1) by comparing the two competing scales with a third, independent scale, calibrated only by the Galaxy, or (2) by comparing with the known value of the basic scale factor of our Galaxy, the galactocentric distance of the Sun, R_0 , with those implied by the distances of the Virgo cluster in the two scales.

The direct calibration of tertiary indicators by means of the Galaxy, without recourse to the galaxies of the Local Group and nearby groups as stepping stones, bypasses most of the difficulties (particularly the propagation of errors) of the classical approach.

Various tests are possible by means of the main metric, photometric and kinematic parameters of the Galaxy (Table 1) which are known with sufficient precision (10–20%) to allow a clear cut choice between two distance scales which differ by a factor of 2.

The morphological type of our Galaxy, SAB(rs)bc, determined by a multiplicity of converging indications^{3,4}, corresponds, in the numerical scale of the *Second Reference Catalogue* (RC2)⁵, to stage $T = 4 \pm 0.5$ of the revised Hubble sequence. The galactocentric distance of the Sun, $R_0 = 8.5 \pm 0.5 \text{ kpc}$, is now established to within 10–15%; it rests on two independent primary distance indicators: the variable stars of the RR Lyr type and those of long periods (Mira-type) whose absolute magnitudes are both fixed by fundamental methods with mean errors of 0.1–0.2 mag (ref. 6). Their agreement greatly strengthens the usefulness of R_0 the basic scale length of the galactic system, as a standard yardstick of extragalactic distances.

The rotation velocity of the Galaxy at the Sun's distance, $V(R_0) = 220 \pm 15 \text{ km s}^{-1}$, is also well determined by several recent investigations⁶. So is the central velocity dispersion $\sigma_v(0) = 130 \pm 7 \text{ km s}^{-1}$ of the galactic bulge⁶. The total absolute magnitude of the Galaxy seen face-on by an external observer, $M_T^0 = -20.2 \pm 0.15$ (in blue light) if $R_0 = 8.5 \text{ kpc}$, can be surprisingly well determined by numerical integration of galactic

models based on star counts and direct photometry of the galactic bulge^{4,7}. (Note that this absolute luminosity scales as R_0^2 , hence M_T^0 varies as $-5 \log R_0/8.5$.) The metric (effective) and photometric (isophotal) diameters of the Galaxy in the standard RC2 system can also be calculated from the photometric models⁴. Finally, several indications suggest that an incomplete ring structure exists in the inner regions of our Galaxy⁴ whose radius is indicated by the '3-kpc arm' feature observed in the 21-cm line of hydrogen out to 20° from the galactic centre⁴, which would give it a diameter of 5.8 kpc.

These basic galactic parameters may be used to calibrate the various correlations between the global properties of galaxies that can serve as distance indicators.

First test: Tully–Fisher relations

The linear correlations discovered by Tully and Fisher (T-F)⁸ between absolute magnitude and maximum rotation velocity V_M (or width W of the 21-cm line) can be used to build extragalactic distance scales. [Traditionally the calibration of these scales (or in practice the zero point of the corresponding distance modulus $\mu = m - M = 5 \log \Delta + 25$, if the distance Δ is in megaparsecs) has been based on a few nearby galaxies whose distances derived from primary or secondary indicators, are presumed to be known. This calibration can be checked now by the zero point independent established by the Galaxy alone.] The most recent version of the T-F relation⁹ for total

Table 1 Fundamental constants of the Galaxy

a, Galactic constants	
Morphological type, SAB(rs)bc	$T = 4 \pm 0.5$
Galactocentric distance of Sun (kpc)	$R_0 = 8.5 \pm 0.5$
Total absolute magnitude (face-on, $i = 0^\circ$)	$M_T^0(B) = -20.2 \pm 0.15$
Total colour index ($i = 0^\circ$)	$(B - V)_T^0 = +0.53 \pm 0.04$
Absolute magnitude of spheroidal component	$M_1^0(B) = -18.2 \pm 0.3$
Colour index of spheroidal component	$(B - V)_1^0 = +0.65 \pm 0.05$
IR absolute magnitude	$M^0(H) = -22.55 \pm 0.23$
Circular rotation velocity (km s^{-1})	$V(R_0) = 220 \pm 15$
Central velocity dispersion (km s^{-1})	$\sigma_v(0) = 130 \pm 7$
Photometric diameter (at $\mu_n = 25.0$)	$D_0/2R_0 = 1.44$
Metric (effective) diameter	$D_e/2R_0 = 0.63$
Diameter of (rs) structure ('3 kpc-arm')	$D_r/2R_0 = 0.34$
b, Stellar constants	
Mean absolute magnitude of RR Lyr variables ($P > 0.42$ days)	$\langle M_v(\text{RR}) \rangle = +0.8 \pm 0.15$
Zero point of bolometric period-luminosity relation of long period variables	$\langle M_{\text{bol}}(0) \rangle = +0.76 \pm 0.1$
Mean absolute magnitude of globular clusters	$\langle M_v(\odot) \rangle = -7.2 \pm 0.2$

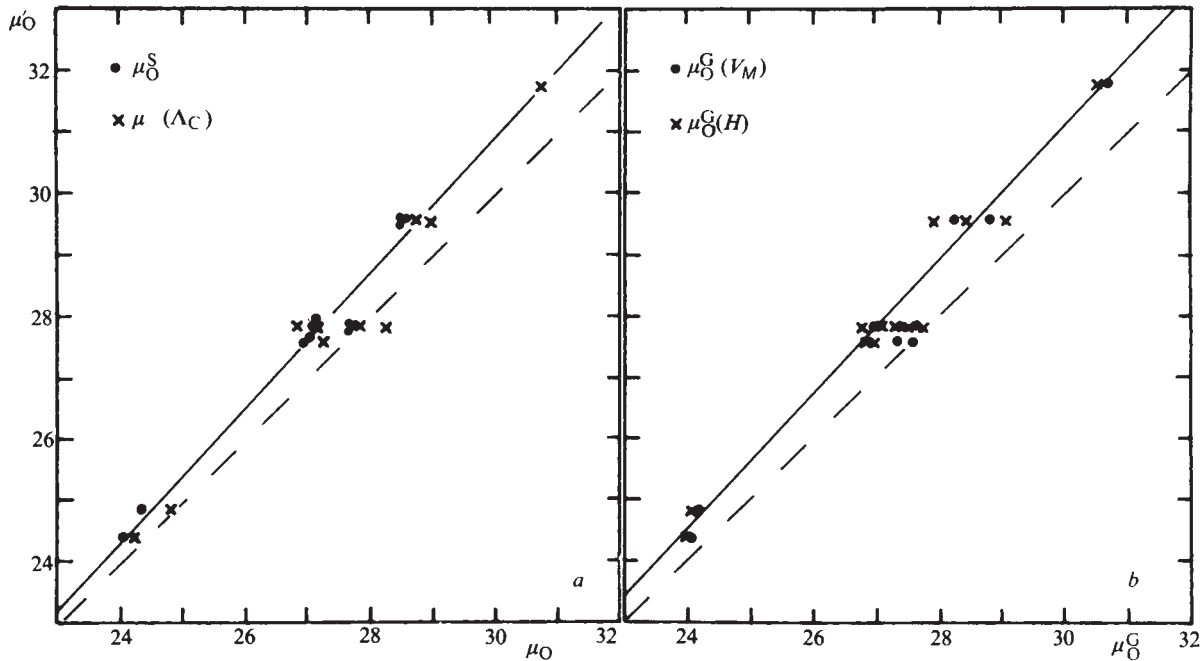


Fig. 1 Comparison of distance moduli on the long distance scale μ'_0 and on several other scales: *a*, primary and secondary calibrators μ_0^S , and luminosity index $\mu_0(\Lambda_c)$ on the short scale; *b*, revised *B*-band T-F relation, $\mu_0(V_M)$ and *H*-band relation, $\mu_0(H)$ with galactic zero points. Both panels show closely similar zero point and scale errors in the long scale. Solid lines, least squares solutions; dashed lines, equality of scales.

magnitudes in the B_T^0 system of RC2 is of the form

$$-M_T^0 = a(B) + 5(\log V_M - 2.2) \quad (1)$$

where V_M (km s⁻¹) is derived from the widths W of the 21-cm line (measured at the 20, 40 and 50% levels of the line profile), corrected for the effects of inclination and velocity dispersion. To calculate the corrected (geometric) distance modulus, $\mu_0 = B_T^0 - M_T^0$, one needs only to fix the zero point, $a(B)$. With the values adopted in Table 1 for the Galaxy, and assuming that $V_M \approx V(R_0)$, as the Sun is on the V = constant branch of the rotation curve⁶, one finds $a^G(B) = 19.5 \pm 0.27$. This value is in excellent agreement with that (19.40 ± 0.15) derived earlier⁹ from 11 calibrating galaxies of the Local Group and nearby groups on the short distance scale. It follows that a correction of -5% only is sufficient to reduce the value of H_0 derived previously¹⁰ from $\mu_0(V_M)$ to the scale whose zero point is fixed solely by the Galaxy, that is $H_0^G = 98 \pm 13$ km s⁻¹ Mpc⁻¹. This value is independent of the supergalactic anisotropy of the velocity field since the value of μ_0 in ref. 10 ($H_0 = 103$) resulted from simultaneous solutions for solar motion and Hubble constant based on 300 galaxies in the distance range $2 < \Delta < 29$ Mpc and well-distributed over both hemispheres.

A second version of the T-F relation applicable to IR magnitudes in the $H^c = H_{0.5}^c$ system¹¹ is of the form

$$-M^c(H) = a(H) + 10(\log W - 2.5) \quad (2)$$

where $W = W_{20}^i$ is the width of the 21-cm line (at 20% level), corrected only for the effect of inclination. To calculate the distance modulus $\mu_0 = H^c - M^c$, one needs only to fix the zero point, $a(H)$. The absolute magnitude $M^c(H)$ of our Galaxy can be computed by correcting $M_T^0(B)$ for the mean value $\langle Y \rangle = 2.35 \pm 0.15$ of the differences $Y = B_T^0 - H^c$ for galaxies of its type T (ref. 12), whence $M^c(H) = -22.55 \pm 0.23$. Then, with

$W = 485$ km s⁻¹, one finds $a^G(H) = 20.69 \pm 0.33$. With this zero point, application of equation (2) to 12 nearby galaxies and to the Virgo cluster (see below) defines a modulus scale (μ_0^G) in excellent agreement (± 0.1 – 0.2 mag) with both the scale based on equation (1) and with the short scale (μ_0), but in strong disagreement with the long scale (μ'_0) (Fig. 1). The mean relations

$$\mu'_0 - 28.18 = (1.133 \pm 0.026)(\mu_0 - 27.61) \quad (3)$$

$$\mu'_0 - 28.18 = (1.120 \pm 0.025)(\mu_0^G - 27.45) \quad (4)$$

confirm the presence of a scale error in the μ'_0 scale, of 12–13% per magnitude in the interval $24 < \mu'_0 < 32$, which had already been suggested by analyses of the foundation and indicators of the long scale².

Application of equations (1) and (2) with their galactic zero points to a sample of 13 spirals of the Virgo cluster (or more precisely the Virgo S Cloud)¹³ leads to mean distance moduli $\langle \mu_0^G(B_T^0, V_M) \rangle = 30.70 \pm 0.12$ (dispersion 0.42 mag) and $\langle \mu_0^G(H, W) \rangle = 30.50 \pm 0.11$ (dispersion 0.40 mag) respectively. This is in good agreement with the short scale: the mean modulus of nine spirals of the Virgo cluster previously derived from their luminosity indices^{13,14} is $\langle \mu_0(\Lambda_c) \rangle = 30.77 \pm 0.16$ (dispersion 0.46 mag). The general mean of a variety of determinations by various methods (all based on the primary and secondary calibrators of the short scale) is $\langle \mu_0 \rangle = 30.71 \pm 0.2$ (refs 2, 15).

The disagreement with the long scale is conspicuous: the distance modulus of the Virgo cluster on the long scale^{1,2} is $\langle \mu'_0 \rangle = 31.5 + 0.26 = 31.76$, where the correction $+0.26$ reflects the revision of the Hyades modulus². If this value were correct, it would imply for the zero point of the scale defined by equation (1), the value $a'(B) = a(B) + (31.76 - 30.70) = 19.5 + 1.06 = 20.56$. In such case equation (1) shows that if $V_0 = 220$ were correct, the absolute magnitude of the Galaxy should be -21.26 (instead of -20.2); and, since M_T^0 scales with $-5 \log R_0/8.5$, it would follow that $R_0^G = 13.8$ kpc (instead of 8.5). Furthermore, the absolute magnitude of the RR Lyr variables should be -0.26 (instead of $+0.8$) and that of the long period variables -0.3 (instead of $+0.76$) which is impossible. Or, if $R_0 = 8.5$

Table 2 Zero points of Faber-Jackson relations

Type	E	L	SO/a	Sa	Sab	Sb	Sbc	Sc	Scd
T	-5	-2	0	1	2	3	4	5	6
$\alpha(B)$	18.98	19.18	20.70	20.88	21.12	21.45	21.88	22.24	23.15
$\alpha(V)$	19.95	19.95	21.52	21.65	21.84	22.12	22.50	23.01	23.67

kpc were correct, one would have to conclude that the rotation velocity of the Galaxy at R_0 is $V_0' = 135 \text{ km s}^{-1}$ (instead of 220), a totally inadmissible value. Or again, if one were to split the corrections between R_0 and V_0 to minimize the errors, one would have to suppose that $R_0' = 10.85 \text{ kpc}$ and $V_0' = 172 \text{ km s}^{-1}$. This combination, which implies errors of 4.2σ in $\log R_0$ and simultaneously of 3.7σ in V_0 , has a total probability (calculated for the normal law) $P(\mu_0') = 1.5 \times 10^{-9}$. The same reasoning applied to equation (2) and to the modulus $\langle \mu_0^G(H) \rangle = 30.53$ leads to similar conclusions.

Second test: Faber–Jackson relations

The linear correlations discovered by Faber and Jackson (F–J)¹⁶ between the absolute magnitude M and the central velocity dispersion $\sigma_v(0)$ in elliptical galaxies or in the spheroidal components of lenticular and spiral galaxies¹⁷ can also be used to build extragalactic distance scales. It is now possible to obtain an independent determination of the zero point by means of the spheroidal component of our Galaxy.

It has been shown recently^{17,18} that the slope, but not necessarily the zero point, of the F–J relation is very nearly the same for galaxies of all three principal classes (E, L, S). The simplest relation is of the form

$$-M_T^0 = \alpha + \beta(\log \sigma_v - 2.3) \quad (5)$$

where M_T^0 is the total magnitude (E, L galaxies) and $\beta(B) = 9$ in the B band, or $\beta(V) = 8.5$ in the V band. In the case of spirals M_T^0 must be replaced by the magnitude M_i^0 of the spheroidal component only. The latter can be deduced from the total magnitude, corrected for the mean systematic difference $\Delta m_i(T) = M_i^0 - M_T^0$, which varies rapidly with morphological type T (ref. 19), and for the colour index $(B - V)_i^0$ of the spheroidal component, which varies also (but little) with T .

For the Galaxy ($T = 4 \pm 0.5$) with $M_T^0(B) = -20.2 \pm 0.15$ (Table 1) and $\Delta m_i(B) = 2.0 \pm 0.1$ (ref. 19), one finds $M_i^0(B) = -18.2 \pm 0.3$ with $(B - V)_i^0 = 0.65 \pm 0.05$. Then with $\sigma_v^G = 130 \pm 7 \text{ km s}^{-1}$ (Table 1) equation (5) gives $\alpha(4, B) = 21.88 \pm 0.40$ (B band) and $\alpha(4, V) = 22.50 \pm 0.42$ (V band). The distance modulus of a galaxy of type T , apparent total magnitude B_T^0 (or V_T^0) and velocity dispersion σ_v is given by

$$\mu_0(B_T^0, \sigma_v) = \alpha(T, B) + B_T^0 + 9(\log \sigma_v - 2.3) \quad (6a)$$

or

$$\mu_0(V_T^0, \sigma_v) = \alpha(T, V) + V_T^0 + 8.5(\log \sigma_v - 2.3) \quad (6b)$$

with the zero points of Table 2. (Note that equations (6) apply to total magnitudes, Δm_i^0 is taken into account in Table 2.)

Table 3 The Galaxy and its 'sospie'

a, Taxonomic parameters								
Parameter	T (1)	L (2)	Λ_c (3)	$(B - V)_T^0$ (4)	ΔM_i (5)	C_{31} (6)	$m_e'(B)$ (7)	H_i (8)
Galaxy	4	2.5:	0.65	0.53	(2.0) [†]	<3.7:	13.14	(2.1) [‡]
Sospie*	4.25	2.25	0.67	0.51	2.25:	3.1:	13.19	2.2
b, Parameters on two distance scales								
Parameter	$M_T^0(B)$ (9)	D_0 (10)	D_e (11)	(12)	O–C (13)			
Galaxy ($R_0 = 8.5 \text{ kpc}$)	-20.2	24.4	10.8	5.8				
Sospie*, if $\langle \mu_0 \rangle = 30.21$	-20.0	22.7	10.3	5.75	+0.2:	+0.1		
—, if $\langle \mu_0 \rangle = 31.73$	-21.5	46.5	21.6	11.4	-1.3:	-1.4		

Columns: (1) type; (2) luminosity class; (3) luminosity index; (4) colour index; (5) relative magnitude of spheroidal component¹⁹ $\Delta M_i = M_i^0 - M_T^0$; (6) concentration index²¹; (7) mean effective surface brightness (mag arc min⁻²) (ref. 5); (8) hydrogen index^{5,19}; (9) total absolute magnitude; (10) photometric diameter (kpc) at isophotal level 25.0 mag arc s⁻²; (11) effective diameter (kpc); (12) diameter of (r_s) ring (kpc), (13) O–C in magnitudes (see text).

* Mean of NGC 1073, 4303, 5921 and 6744.

† Mean for $T = 4$.

‡ Mean for $\Lambda_c = 0.65$.

Table 4 Implications of two estimates of the distance to the Virgo E cluster

Parameters	Long scale ($H_0 = 50$)	Short scale ($H_0 = 95$)	Observed
Distance of Virgo E cluster	22.5 (E, S)	11.8 (E)	
True distance modulus μ_0	31.76 (μ_0)	30.37 (μ_0)	
Galactic extinction A_v	0.00	0.15	
Apparent modulus μ_v	31.76	30.52	
Mean apparent magnitude of globular clusters	$\langle m_v \rangle$ m.e.		23.3 ± 0.04
Mean absolute magnitude of globular clusters	$\langle M_v(\oplus) \rangle$ m.e.	-8.46 -7.22	-7.2 ± 0.2
Mean absolute magnitude of RR Lyr variables	$\langle M_v(RR) \rangle$ m.e.	-0.46 +0.78	+0.8 ± 0.15
Distance of galactic centre	R_0 (kpc) m.e.	15.2 ± 3.1	8.4 ± 0.5
Zero point of P–L relation of Mira variables	$\langle M_{bol}(0) \rangle$ m.e.	-0.5 ± 0.4	+0.74 ± 0.1

The validity of these relations has been verified by application to four nearby spirals (M31, M81, NGC 4594 and 7793) covering a wide range of types and a large interval of σ_v . Equations (6) were then applied to a sample of 15 spirals having velocity dispersions $130 < \sigma_v < 180 \text{ km s}^{-1}$. The distance moduli so calculated were compared with those previously derived, on the short scale, from primary and secondary indicators, μ_0^G , from the luminosity index, $\mu_0(\Lambda_c)$ (ref. 14), from the T–F relations in the B -band, $\mu_0(V_M)$ (ref. 9), and from the diameter of the r , r_s inner rings, $\mu_0(r)^2$. There is excellent agreement ($\pm 0.2 \text{ mag}$) and no appreciable divergence of the scales in the interval $24 < \mu_0 < 32$ covered by the data. The mean systematic difference $\langle \mu_0^G(\sigma_v) - \mu_0 \rangle = +0.10 \pm 0.17$ is insignificant and the dispersion of the differences (0.78 mag) is consistent with the estimated errors (0.5–0.6 mag) of the various indicators.

A comparison with the moduli (μ_0^G) of the same galaxies on the long scale leads to very different results. In the interval $24 < \mu_0 < 32$, the mean ratio is $(\mu_0 - 30.07)/(\mu_0^G(\sigma_v) - 29.22) = 1.088 \pm 0.084$ (Fig. 2b), which confirms the presence of zero point and scale errors in the μ_0^G moduli, in agreement with the conclusions drawn from equations (3) and (4).

To apply equations (6) to the galaxies of the Virgo E cluster²⁰ the differences of zero points reflecting the fundamental differences in shape and internal dynamics between elliptical galaxies and the bulges of spirals must be taken into account. At $\sigma_v = \text{constant}$, ellipticals are on average less luminous than the bulges of spirals by $\sim 0.7 \text{ mag}$ according to Whitmore *et al.*¹⁷ which corresponds to 0.9 mag in the system of Simien and de Vaucouleurs¹⁹; it follows that $\alpha = 18.98$ (B) and 19.95 (V) for $T = -5$. These constants are subject to revision because the magnitude difference at $\sigma_v = \text{constant}$ between elliptical and spiral bulges is still in doubt, mainly as a result of different methods in decomposing spirals into disk and spheroidal components. Hence the present application of the F–J relations to the Virgo E cluster carries a lower weight than the other tests. Rejecting it would make no difference to the conclusions (see Table 5). According to de Vaucouleurs and Olson¹⁸ elliptical and lenticular galaxies are equally luminous (at $\sigma_v = \text{constant}$) in the V_T^0 system of total magnitudes, but lenticulars are brighter by 0.2 mag in the B_T^0 system. This implies $\alpha = 19.18$ (B) and 19.95 (V) for $-3 \leq T \leq -1$ (Table 2).

With the zero points so fixed (indirectly) by the Galaxy equations (6) have been applied to 17 early-type galaxies ($T < 0$) of the Virgo E cluster²⁰. The mean modulus $\langle \mu_0^G(\sigma_v) \rangle = 30.32 \pm 0.10$ (internal m.e.) is in excellent agreement with the mean value $\langle \mu_0 \rangle = 30.37 \pm 0.2$, previously derived from a variety of indicators on the short scale^{2,15}.

Both values are in sharp disagreement with the modulus of the Virgo cluster, $\mu_0' = 31.76$ (with the corrected Hyades zero point)^{1,2} postulated by the long scale. The discrepancy $\langle \mu_0' - \mu_0^G(\sigma_v) \rangle = +1.44 \text{ mag}$ is far in excess of any plausible error in the galactic calibration of the σ_v scale. A propagation of error analysis shows that if the difference were distributed among the four parameters R_0 , T , σ_v , and the correction $\Delta M(E-S)$, so

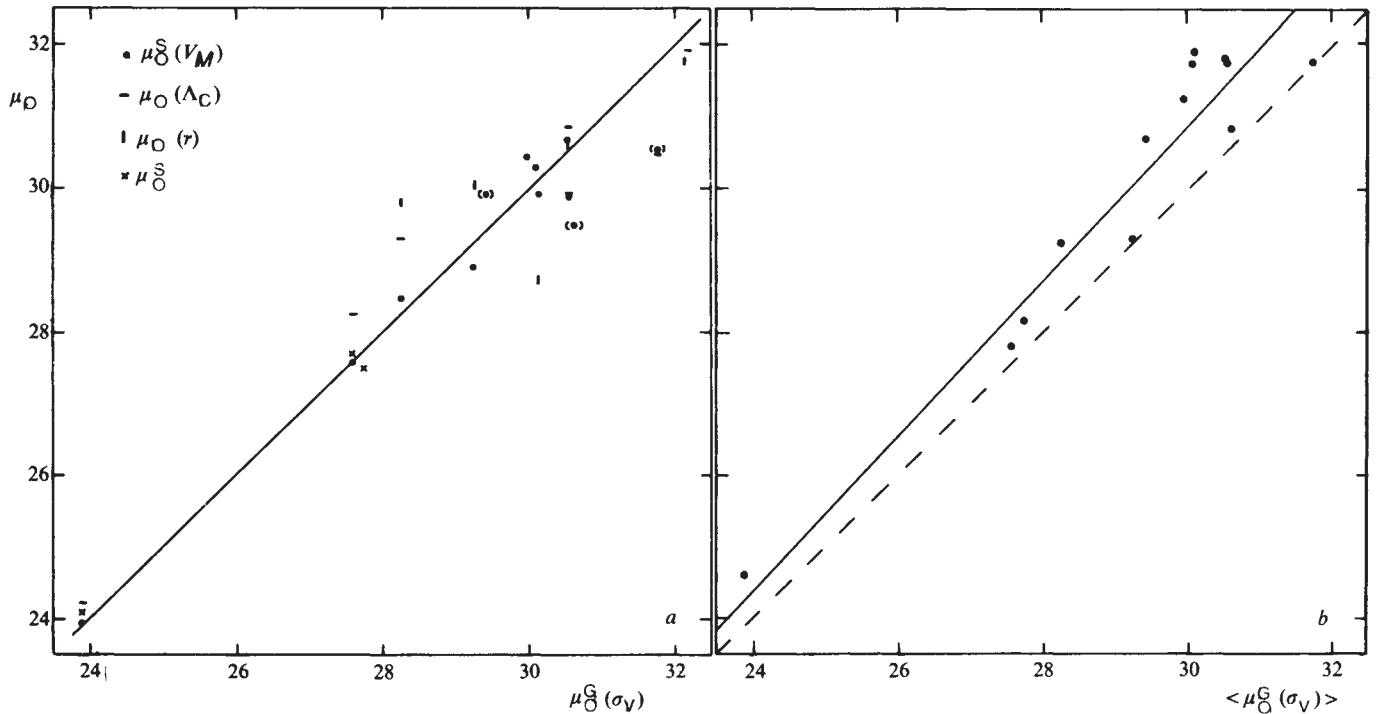


Fig. 2 Comparison of distance moduli: *a*, on the short scale (see text); *b*, on the long scale, with the mean values $\mu_0^G(\sigma_V)$ derived from the F-J relations. Note good systematic agreement in *a* and zero point and scale error in *b*.

as to minimize the errors, a reconciliation of the long scale with the galactic parameters would require changing simultaneously (and in the right direction) each of the four parameters by more than 1.4σ . This combination of errors has a total probability $P(\mu_0') < 10^{-5}$.

Third test: zero point of the luminosity index scale

The tertiary index on which the short distance scale was initially based^{2,14} is the corrected 'luminosity index', $\Lambda_c = (T + L_c)/10$, where T is the morphological type and L_c the luminosity class, corrected for inclination effects¹⁴. In the interval $|\Lambda_c - 1| < 0.65$, the correlations between Λ_c and the total absolute magnitude, M_T^* , or the linear photometric (isophotal) diameter, D_0 , both in the RC2 system⁵, are essentially linear:

$$-M_T^* = \gamma - 3.0(\Lambda_c - 1) \quad (7)$$

and

$$\log D_0 = \delta - 0.6(\Lambda_c - 1) \quad (8)$$

The constants γ, δ fixing the zero points of the distance scales were first determined by means of 16 'calibrating galaxies', that is, nearby galaxies whose distances had been previously derived from primary and secondary indicators^{2,15}. The slopes of the relations are well determined and the linearity of the scale of distances based on Λ_c has already been verified by comparison with a variety of independent indicators^{2,9,15}. It remained to confirm the adopted values of γ, δ by direct application of equations (7) and (8) to our Galaxy whose type, Sbc ($T = 4 \pm 0.5$), is well established. Its luminosity index is not directly observable, but can be inferred indirectly by three methods: (1) the statistical relation $\langle T \rangle = \langle L \rangle + 1$ (ref. 21) suggests $L^G \approx 3$ and $\Lambda_c^G \approx 0.7 \pm 0.1$; (2) according to all available evidence^{3,4} our Galaxy is intermediate between M31 ($T = 3, L = 2, \Lambda_c = 0.45$) and M33 ($T = 6, L = 4, \Lambda_c = 1.00$) (ref. 14), but closer to the former, whence $\Lambda_c^G \approx 0.63 \pm 0.07$; (3) the statistical relation²¹ between Λ and the maximum rotation velocity V_M , log

$V_M = 2.15 - \frac{1}{2}(\Lambda - 1)$, implies $\Lambda_c^G \approx 0.61 \pm 0.1$, if $V_M \approx V_0 = 220 \pm 15 \text{ km s}^{-1}$. On average $\Lambda_c^G = 0.65 \pm 0.05$. Then, with the adopted values of $M_T^* = -20.2 \pm 0.15$ and $\log D_0 = 4.39 \pm 0.03$ (D_0 in parsecs) for $R_0 = 8.5 \text{ kpc}$ (Table 1), equations (7) and (8) give $\gamma^G = 19.15 \pm 0.25$ and $\delta^G = 4.18 \pm 0.05$.

These values, which happen to coincide with those, $\gamma = 19.15 \pm 0.16$ and $\delta = 4.18 \pm 0.045$, previously derived from the 16 calibrating galaxies¹⁴, confirm precisely the linear formulae adopted to calculate the distances of 458 galaxies¹⁴ on the short scale and the value of the Hubble constant $H_0 = 96 \pm 10 \text{ km s}^{-1} \text{ Mpc}^{-1}$ derived from them²². [Actually two different forms of interpolating formulae were used in ref. 14, and the resulting weighted mean moduli $\mu_0^G(\Lambda_c)$ are -0.07 less than those calculated from equations (7) and (8) with the γ^G, δ^G zero points; hence, with the latter, distances would be 3.2% greater and H_0 3.2% smaller, that is $H_0 = 93 \pm 11 \text{ km s}^{-1} \text{ Mpc}^{-1}$.] This value is independent of the supergalactic anisotropy of the velocity field because the value of μ_0 in ref. 22 ($H_0 = 96$) resulted from simultaneous solutions for solar motion and Hubble constant based on 200 galaxies in the distance range $2 < \Delta < 32 \text{ Mpc}$ and well-distributed over both hemispheres.

Clearly the probability that $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ is very small calculated for a one-tail normal error law in log H_0 it is $P(\mu_0') < 10^{-7}$.

Table 5 Probabilities $P(\mu_0')$ that the fundamental constants of the Galaxy are consistent with the long distance scale

Tests	Parameters	$P(\mu_0')$
Tully-Fisher relation in Virgo S cluster	$R_0, V_0, M_T^* B\text{-band}$	1.5×10^{-9}
Faber-Jackson relations in Virgo E cluster	$R_0, \sigma_v(0), M_T^*, T$	2×10^{-6}
Magnitude-diameter-luminosity index relations for 458 spirals	$R_0, M_T^*, D_0, \Lambda_c$	$< 10^{-7}$
Magnitude and diameters of Galaxy compared with its 'sosie'	$R_0, M_T^*, D_0, D(r)$	$< 10^{-8}$
Mean magnitudes of globular clusters in Virgo E cluster	$R_0, M_T^*, m_v(\oplus), M_v(RR)$ $M_{bol}(0) \text{ (Miras)}$	$< 10^{-5}$

Fourth test: Galaxy and its 'sosies'

The concept of galactic 'sosies' has recently been introduced by Paturel²³ to describe galaxies that look essentially identical with respect to all available taxonomic criteria (such as type, colours, bulge-to-disk ratio, maximum rotation velocity and hydrogen-luminosity ratio). By application of Hubble's 'principle of the uniformity of nature', it is reasonable to expect them to have also closely similar luminosities and sizes.

According to numerous indices the four galaxies NGC 1073, 4303, 5921 and 6744, appear closely to resemble our Galaxy^{4,24}. The composite galaxy defined by the average of these four galaxies is a good 'sosite' of our Galaxy (Table 3a). On this basis the plausibility of proposed distance scales can be tested by comparison of the known values of the absolute magnitude and linear diameters of our Galaxy (Table 1) with those calculated for its 'sosite' in each of the two following cases: (1) $\langle\mu_0\rangle \equiv \langle\mu_0^w(\Lambda_0)\rangle = 30.21$ on the short scale and corresponding to $H_0 = 95 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (ref. 22), and (2) $\langle\mu_0\rangle = 31.73$ on the long scale (with corrected Hyades zero point) and assuming $H_0 = 50$ (ref. 25). Table 3b shows that on the average the 'sosite' appears only 0.2 mag fainter than our Galaxy on the short scale, but 1.3 mag brighter on the long scale. Likewise, the diameters of the 'sosite' are $\sim 4\%$ less (corresponding to 0.1 mag fainter) on the short scale, but $\sim 96\%$ larger (corresponding to ~ 1.4 mag brighter) on the long scale. It is obvious that while the short scale implies a quasi-equality between the Galaxy and its 'sosite', as suggested by the taxonomic criteria, the long scale leads to a contradiction. To reconcile the metric and photometric parameters of the Galaxy and its 'sosite' on the long scale, one would need to change R_0 from 8.5 to ~ 16 kpc which would require a 4.5σ change in M_T . It would also imply that the mean absolute magnitudes of the RR Lyr and long period variables are both 1.3–1.4 mag brighter than the best current estimates (Table 1), which is out of the question.

Fifth test: distance of the Virgo cluster and its implications

A final verification of the short scale and an independent confirmation of the impossibility of the long scale rest on a comparison of the implications of the distances proposed for the Virgo cluster, particularly the Virgo E cluster, in which many globular star clusters associated with spheroidal galaxies have been observed. According to Hanes²⁶ the mean apparent magnitude of these clusters, derived from their luminosity function, on the plausible assumption that it is gaussian with dispersion $\sigma_M = 1.1 \pm 0.1$ mag (ref. 27), is $\langle m_v \rangle = 23.3 \pm 0.4$. As shown in Table 4 this implies $\langle M_v(\oplus) \rangle = -7.22$, if Virgo E is at the distance (11.8 Mpc) corresponding to the mean distance modulus $\langle\mu_0\rangle = 30.37 \pm 0.2$ on the short scale, but $\langle M_v(\oplus) \rangle = -8.46$, if it is at the distance (22.5 Mpc) corresponding to the mean modulus $\langle\mu_0\rangle = 31.76$ postulated by the long scale^{1,2}.

Now, the mean magnitude difference between the globular clusters and the RR Lyr variables is $\langle M_v(\text{RR}) - M_v(\oplus) \rangle = 8.0 \pm 0.1$ (ref. 15). It follows that the short scale implies $\langle M_v(\text{RR})(\mu_0) \rangle = -7.22 + 8.0 = +0.78 \pm 0.4$, in excellent agree-

ment with the average of modern determinations of this primary indicator, $\langle M_v(\text{RR}) \rangle = +0.8 \pm 0.15$ (Table 1). But the long scale implies $\langle M_v(\text{RR})(\mu_0) \rangle = -8.46 + 8.0 = -0.46 \pm 0.4$, which is $1.26 \text{ mag} \approx 3\sigma$ brighter than the adopted value.

Furthermore, because the values of R_0 derived from RR Lyr and Mira-type variables are in excellent agreement (Table 1)⁶, one would also have to conclude that the zero point of the (bolometric) period–luminosity relation of the long period variables should be changed from $+0.76$ to -0.5 ± 0.4 , which would be another 3σ correction. The combined probability of such errors, that is of the hypothesis that the long scale is correct, is $P(\mu_0) < 2 \times 10^{-6}$.

Discussion and conclusions

The results of the five tests above, summarized in Table 5, singly and collectively demonstrate that reconciling the long scale with the known values of the basic metric, photometric and kinematic parameters of the Galaxy is practically impossible.

A detailed discussion of the various *ad hoc* hypotheses that might be advanced to escape this conclusion and 'save' the long scale (a different galactic extinction model, a large cosmic scatter about the T–F and F–J relations, large errors in the adopted values of the galactic parameters, systematic differences between the globular cluster luminosity functions in the Virgo cluster and the Galaxy) shows that none is plausible nor capable of quantitatively accounting for the 1.1–1.4 mag discrepancies. The conclusion seems inescapable that first on the short scale all galactic data fit in naturally within their known errors and there is no need to invoke special circumstances, and second the long scale fails all available tests by a large margin and cannot be reconciled with the galactic data.

The essentially independent scales based on the traditional approach, that is on primary, secondary and tertiary indicators², and the new approach using the Galaxy alone as a fundamental calibrator, are in near perfect agreement and the most probable value of the Hubble constant derived from them is $H_0 = 95 \text{ km s}^{-1} \text{ Mpc}^{-1}$.

The time scale argument often invoked in support of $H_0' = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and as an objection to the Hubble time, $T_H = H_0^{-1} \approx 10^{10} \text{ yr}$, associated with $H_0 = 95 \text{ km s}^{-1} \text{ Mpc}^{-1}$, rests entirely on the arbitrary (often implicit) hypothesis that a Friedmann model without cosmological constant is applicable to the real Universe. It is interesting that recent studies of the space distribution of quasars²⁸ lead, independently of any assumption on H_0 , to a Lemaître model with a small positive curvature parameter ($k = K/H_0^2 = +0.22$), a reasonable density parameter ($\Omega_0 \approx 0.07$) and a well-determined positive cosmological constant ($\Lambda = \Lambda/3H_0^2 = +1.15$), which implies a satisfactory value for the age of the Universe, $T_U \approx 1.8 \times 10^{10} \text{ yr}$ if $T_H = 10^{10} \text{ yr}$.

This work was supported by a visiting fellowship of the Australian National University while G. de V. was on research leave from the University of Texas. I acknowledge the support of both institutions and the hospitality of Mount Stromlo and Siding Spring Observatories.

Received 10 May; accepted 24 June 1982.

1. Tammann, G. A., Sandage, A. & Yahil, A. in *Physical Cosmology*, Les Houches, Sess. XXXII, 53 (North Holland, Amsterdam, 1979).
2. de Vaucouleurs, G. in *10th Texas Symp. in Relativistic Astrophysics*, 90–122 (New York Academy of Sciences, 1981).
3. van den Burgh, S. *The Galaxies of the Local Group* (Preprint, David Dunlap Observatory, No. 195, 1968).
4. de Vaucouleurs, G. & Pence, W. D. *Astr. J.* **83**, 1163–1173 (1978).
5. de Vaucouleurs, G., de Vaucouleurs, A. & Corwin, H. G. *Second Reference Catalogue of Bright Galaxies* (University of Texas Press, Austin, 1976).
6. de Vaucouleurs, G. *Astrophys. J.* (submitted).
7. Bahcall, J. & Soneira, R. M. *Astrophys. J. Suppl.* **44**, 73–110 (1980).
8. Tully, R. B. & Fisher, R. *Astr. Astrophys.* **54**, 661–673 (1977).
9. Bottinelli, L., Gouguenheim, L., Paturel, G. & de Vaucouleurs, G. *Astrophys. J. Lett.* **242**, L153–L156 (1980).
10. de Vaucouleurs, G., Peters, W. L., Bottinelli, L., Gouguenheim, L. & Paturel, G. *Astrophys. J.* **248**, 408–422 (1981).
11. Aaronson, M., Mould, J. & Huchra, J. *Astrophys. J.* **229**, 1–13 (1979); **237**, 655–665 (1980).
12. Tully, R. B., Mould, J. & Aaronson, M. *Astrophys. J.* (in the press).

13. de Vaucouleurs, G. *Astrophys. J.* **253**, 520–525 (1982).
14. de Vaucouleurs, G. *Astrophys. J.* **227**, 380–390, 729–755 (1979).
15. de Vaucouleurs, G. *Observatory* (in the press).
16. Faber, S. & Jackson, R. *Astrophys. J.* **204**, 668–683 (1976).
17. Whitmore, B. C., Kirshner, R. P. & Schechter, P. L. *Astrophys. J.* **234**, 68–75 (1979); **250**, 43–54 (1981).
18. de Vaucouleurs, G. & Olson, D. W. *Astrophys. J.* **256**, 346–369 (1982).
19. Simien, F. & de Vaucouleurs, G. *IAU Symp.* No. 100 (submitted).
20. de Vaucouleurs, G. *Astrophys. J. Suppl.* **6**, 213–234 (1961).
21. de Vaucouleurs, G. *The Evolution of Galaxies and Stellar Populations*, 43–94 (Yale University Observatory, New Haven, 1977); *Astr. J.* **85**, 182 (1980).
22. de Vaucouleurs, G. & Peters, W. L. *Astrophys. J.* **248**, 395–407 (1981).
23. Paturel, G. thesis, Univ. Lyon (1981); *Astr. Astrophys.* (submitted).
24. de Vaucouleurs, G. *IAU Symp.* No. 38, 18–25 (1970).
25. Sandage, A. & Tammann, G. A. *A Revised Shapley-Ames Catalogue of Bright Galaxies* (Carnegie Institution Publ. 635, Washington, 1981).
26. Hanes, D. A. *Mon. Not. R. astr. Soc.* **180**, 309–321 (1977); **188**, 901–909 (1979).
27. de Vaucouleurs, G. *Nature* **266**, 126–129 (1977).
28. Fliche, H. H. & Souriau, J. M. *Astr. Astrophys.* **78**, 87 (1979).

Time-resolved X-ray diffraction studies of the structural behaviour of myosin heads in a living contracting unstriated muscle

J. Lowy & F. R. Poulsen

Open University, Oxford Research Unit, Foxcombe Hall, Boars Hill, Oxford OX1 5HR, UK

The intensities of three regions of the low-angle X-ray diffraction pattern from a molluscan unstriated muscle have been followed during tension generation at a time resolution of 0.5–1 s using synchrotron radiation. The observed intensity changes can be reasonably interpreted in terms of myosin cross-bridge movements during the contractile cycle. A model that accounts for the intensity changes suggests that myosin heads move out from the thick filament during activation and attach to actin sites to produce tension with a small delay. During relaxation from both phasic and tonic contractions the heads remain attached to actin sites longer than it takes for tension to decay.

THERE is now general agreement that the transformation of chemical to mechanical energy in muscle is associated with a calcium-controlled interaction between actin and myosin during which ATP is split. The globular head of the myosin molecule is believed to attach and detach from actin in a cyclical manner and this is called the 'cross-bridge cycle'^{1,2}. We have used X-ray diffraction techniques to obtain time-resolved data about the behaviour of myosin heads during the mechanical contractile cycle in an unstriated muscle, the anterior byssus retractor (ABRM) of *Mytilus edulis*. Such investigations aim to produce structural information from normally functioning living muscles which is not only interesting for its own sake but can also help in the interpretation of biochemical data and formulation of kinetic schemes. The choice of the ABRM as our experimental material was motivated by the fact that it possesses certain unique structural and mechanical properties.

We define a contractile unit in the ABRM as the thick filament with its surrounding actin filaments. The unique structural feature of this muscle is that its thick filament diameter shows a wide variation^{3,4} which means that its contractile units cannot form a well-ordered lattice. This feature is also responsible for the smooth, streak-like appearance of the central equatorial diffraction (Fig. 1a).

Previous X-ray studies of the ABRM⁵ have demonstrated that the sideways packing of actin filaments produces a strong equatorial peak (Fig. 1a) at 14 nm. Our present interpretation of changes in the intensity of the peak is based on the assumption that the actin filaments are packed into a one-dimensional lattice (rosette) which surrounds the thick filament. This is supported by evidence from electron micrographs from this laboratory and by numerical modelling of the intensity distribution of the equatorial peak.

In addition to diffraction from the thick and thin filaments, we describe here for the first time the diffuse background (disk) diffraction (Fig. 1a) and show that this is likely to be associated with disordered myosin heads.

Stimulation of the ABRM with alternating current produces a phasic contraction⁶. After stimulation stops tension decays to zero somewhat more slowly than tension rise⁶. Treatment with acetylcholine⁷ gives a tonic response known as a catch contraction. After washing out the drug, relaxation can be up to two orders of magnitude slower than in a phasic contraction. From the physiological viewpoint the muscle is unique in that in catch contractions a high level of tension is maintained in the absence of active state (that is, ability for redevelopment of tension after a release^{8,9}) and energy turnover as measured by oxygen consumption¹⁰ is so low that the amount of ATP split must be insignificant.

Another drug, serotonin, exerts its most important effect on relaxation⁷. Thus, with increasing concentration, serotonin

speeds up tension decay in both phasic and catch contractions^{7,8}, until the rate reaches a value nearly as high as tension rise⁸. With natural seawater as a bathing solution, different muscles produce contractions with different relaxation rates. To obtain standardized muscles, serotonin was applied at a concentration of $10^{-4.5}$ M, which produces a completely reproducible relaxation rate¹¹.

In a standardized ABRM the half time of tension rise is ~ 1.5 s at 10°C , about 50 times slower than the corresponding time at the same temperature for twitches of the striated frog's sartorius used by Huxley *et al.* in similar time-resolved X-ray experiments¹². In phasic contractions we were able to obtain useful data with a time resolution of 0.5–1 s. A time resolution of 1–3 min was adequate for the relaxation phase in catch contractions.

The preparation of the ABRM has been described previously, as have optimal methods of stimulation¹¹. The muscles used were about 25 mm long and 1 mm thick. They were maintained in oxygenated natural seawater (osmolality about 600 mosmol) in a Perspex cell with mylar windows for the passage of X rays. The cell was fitted into a copper frame through which liquid from a heating/cooling bath was circulated in order to control the temperature of the muscle's bathing solution. Muscles were mounted isometrically with one end attached to a force transducer.

The experiments were carried out at the EMBL Outstation at DESY, Hamburg, using the electron-positron storage ring DORIS as the X-ray source. Both tension and X-ray data were transmitted into a computer. Details of the Hamburg set-up are given elsewhere¹³.

All our results were obtained with a muscle-to-detector distance of 3 m. Phasic contractions were produced by applying a.c. (50 Hz) via transverse electrodes for 6 s and an interval of 3 min was allowed between contractions¹¹. Three features of the low-angle X-ray diffraction pattern were studied using two placements of a position-sensitive detector (Fig. 1a, at P and Q). The X-ray features are the equatorial streak, the 14-nm equatorial peak and the central diffuse disk (Fig. 1a). All subsequent references to streak intensity mean the intensity at the 38-nm equatorial spacing (Fig. 1a, at Q).

Results

The equatorial streak and central diffuse disk were recorded together by placing the detector in the meridional direction at an equatorial spacing of 38 nm (Fig. 1a, at Q). The resting intensity profile thus obtained is shown in Fig. 2a together with the difference profile at peak tension. The boundary between streak and disk is indicated by the double arrows. At peak tension the streak has increased and the disk decreased in

intensity. Figure 3a shows the time course of the streak intensity integrated after subtraction of the disk-background. The intensity increases with an initial rate appreciably higher than that of tension which itself follows a sigmoidal time course. This gives the streak a 1.5-s lead in reaching half-maximum. About 3 s ahead of tension it reaches a plateau 25% above its resting intensity where it stays for about 12 s before returning with a 5-s lag behind tension at half-maximum. The intensity change recorded in the central disk (Fig. 3a) shows a decrease of about 35% which follows much the same time course as tension rise; it returns to its resting value more slowly than tension.

Two other ways of presenting the disk data are useful. First, the time dimension can be eliminated by plotting disk intensity against tension. This reveals a linear relationship during the rising phase but not during relaxation (Fig. 4a). It also allows pooling of results from experiments with different time courses. Second, to determine the relative number of scattering units

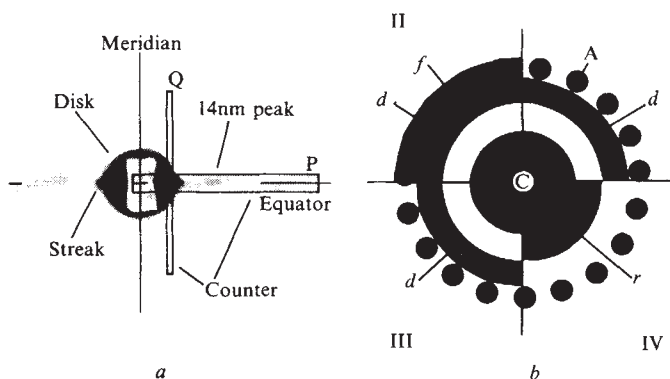


Fig. 1 *a*, Low-angle X-ray diffraction pattern of resting ABRM muscle in natural seawater recorded on film with an exposure time of 2.5 min and a muscle-to-film distance of 3 m. Note the equatorial streak, central diffuse disk and 14-nm equatorial peak. The two positions (P and Q) of the one-dimensional position-sensitive detector are shown in terms of the effective window used in the time-resolved experiments with contracting muscles. Detector position Q is at an equatorial spacing of 38 nm. The boundary between the streak and the 14 nm peak is at an equatorial spacing of about 20 nm. *b*, Model of contractile unit used for the interpretation of equatorial intensity changes. Division into four quadrants represents phases in the mechanical contractile cycle which are referred to in the text. In the ABRM a variety of thick filament diameters are present³ and the frequency of occurrence f_i has been determined⁴ ($\sum f_i = 1$). The paramyosin core C of radius r_{Ci} is surrounded by a rosette of $N_{Ai} = 2\pi r_{Ai}/14$ nm evenly spaced actin filaments A (centre-to-centre distance 14 nm) of 8 nm diameter and at a radial position $r_{Ai} + r_{Ci} + 24$ nm. The radial position r_{sil} of the centres of mass of the myosin heads can vary from $r_{Ci} + 5$ nm at r , to $r_{Ci} + 15$ nm at d , to $r_{Ci} + 24$ nm at f . The intensity of the equatorial streak is $\sum f_i \cdot I_i$, where I_i is the intensity from the i th size of unit: $I_i = (F_{Ci} + F_{Ai} + F_{Si})^2$, where $F_{Ci} = w_{Ci} \cdot 2J_1(2\pi r_{Ci}R)/(2\pi r_{Ci}R)$, $F_{Ai} = w_{Ai} \cdot J_0(2\pi r_{Ai}R)$ and $F_{Si} = w_{Si} \cdot J_0(2\pi r_{Si}R)$. These formulae are explained by Vainshtein²⁴. w_{Ci} , w_{Ai} and w_{Si} are the respective masses of paramyosin, actin plus tropomyosin, and myosin head in a 72-nm long section of a contractile unit of the i th size. The equatorial position R is $1/38$ nm. At this value of R higher order Bessel functions can be omitted in the expression for F_{Ai} as can the contribution from the shape factors for the various molecules. The calculations show that if all myosin heads move from r to d , the streak intensity at 38 nm increases by 11% and that if they move to f , the increase is 46%. The 14-nm peak comes basically from the evenly spaced actin filaments. Its intensity is reduced as myosin heads enter the spaces between actin filaments. In our model practically all actin filaments are part of the rosettes, and this, together with our two-state structural model (see text), means that the intensity of the 14-nm peak is of the following form: $I = I_0 \cdot (\mu_A - x\mu_{Si})^2$, where I_0 is the intensity with no myosin heads in the rosette spaces, μ_A and μ_{Si} are, respectively, the total masses of actin filaments and of myosin heads, and x is the fraction of the myosin heads present in the rosette spaces. As shown in the text, μ_{Si}/μ_A is 0.29. This means that the maximum possible intensity decrease is to 50%. Although the above expression for the intensity as a function of number of myosin heads in the rosette spaces is parabolic, the relation is practically linear over the range where the intensity is reduced by a factor of two. In our model the myosin heads are disordered in locations r and d . They produce the central diffuse disk which will have an intensity directly proportional to the number of disordered heads. The intensity of the disk will be independent of the radial position of the myosin heads, but as they enter the rosette spaces, they will become sufficiently ordered not to contribute to the disk.

(see below) the intensity data can be drawn as Guinier plots¹⁴ (Fig. 5). To record the 14-nm equatorial peak the detector was oriented in the horizontal direction and placed as shown at P in Fig. 1a. The intensity profile thus obtained at rest and peak tension is seen in Fig. 2b.

During the rising phase, the intensity of the 14-nm peak decreases by maximally 50% with much the same time course as tension; it returns to its resting value appreciably more slowly than tension (Fig. 3b). The 14-nm peak does not broaden in the equatorial direction (Fig. 2b).

Plotting intensity against tension gives a linear relation during the rising phase but not during relaxation (Fig. 4b). Except for a difference in time scale the same results were obtained in catch contractions (Fig. 4c).

Structural interpretation of intensity changes

We consider models for the contractile units like that shown in Fig. 1b and take into account the known variation in their size. We assume constancy in the axial projection of the structure of the paramyosin core and the location of the actin filaments. Hence, the change in the location of the myosin heads will account for all the intensity changes. The reason for assuming constancy in the paramyosin core structure is that we observe no changes in intensity of the 36-nm paramyosin layer line during contraction. Arguments for a constant actin filament location are given in the discussion of the 14-nm equatorial peak.

Quantitative interpretation of the equatorial intensity changes is greatly assisted by knowing the relative masses of the proteins involved. To this end we calculate the masses of the proteins in a 72-nm long section for the various sizes of the contractile units. It is assumed that the number n_p of paramyosin molecules per nm² of cross-sectional area and the number n_m of myosin molecules per nm² of surface area of the core is independent of the core diameter, and that n_p and n_m are the same in the ABRM and in the opaque adductor of the oyster which has thick filaments of a similar type¹⁵. For the opaque adductor the frequency of occurrence f_i (expressed as a fraction of the total) of core diameter D_i has been determined using electron microscopy¹⁶. We calculate n_p to be 0.184 from X-ray data on the intermolecular distance extrapolated to the dehydrated state of the muscle^{17,18} (tetragonal unit cell side taken as 1.65 nm), and from the molecular repeat length of 144 nm (ref. 19). Taking the molecular weights (M_r s, in daltons) for paramyosin (220,000) and myosin (470,000), and 8 as the mass ratio paramyosin/myosin for the opaque adductor²⁰, we find that n_m comes to 0.00302. We can now calculate the paramyosin and myosin masses in a contractile unit of length 72 nm as a function of core diameter.

We also want to know the average masses per 72 nm in the ABRM. Using f_i from ref. 4, we find that these values for paramyosin and myosin are 5.670 and 1.277×10^7 daltons, respectively. Then, taking the M_r of the myosin head as 121,000 (ref. 21), the myosin head mass comes to 0.658×10^7 daltons.

Taking the M_r s of actin (42,000) and tropomyosin (70,000), a 2.75-nm axial repeat between actin monomers, and 7 actin monomers per tropomyosin molecule, we calculate that the mass m_A in a 72-nm long actin filament is 1.361×10^6 daltons. The number of actin filaments N_{Ai} surrounding a given thick filament can be determined from the core diameter (see Fig. 1b) and the spacing of 14 nm between actin filaments. However, due to the sharing of actin filaments between thick filaments, only the fraction $s = N_0/\sum f_i \cdot N_{Ai}$ of their mass belongs to the contractile unit. N_0 is the observed average number of actin filaments per thick filament. Knowing that $N_0 = 16.7$ (ref. 4), we find that $s = 0.848$. The average mass of actin filament per 72 nm length is $N_0 \times m_A = 2.274 \times 10^7$ daltons. Thus, the mass ratio myosin head/actin filament comes to 0.29.

Taking into account the variation in the size of the contractile units, our calculations show that if all heads move from r to d in Fig. 1b, the average streak intensity at 38 nm increases by 11%; if they move to f the increase is 46%. In our experiments

the actual intensity increase at peak tension ranges from 25 to 50%. This indicates that during tension rise myosin heads move radially outward from the thick filament surface and at least partly into the rosette spaces.

Our finding that during contraction the intensity decrease in the 14-nm peak is not accompanied by any change in its equatorial width (Fig. 2b) means that the intensity decrease is

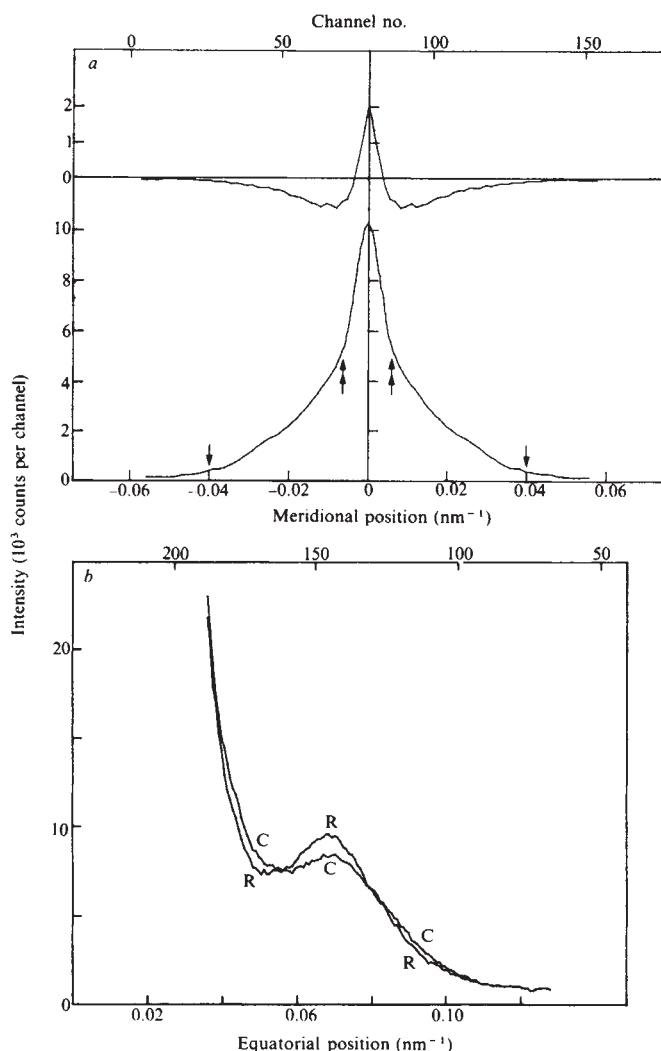


Fig. 2 *a*, Intensity profiles obtained with the detector in the meridional direction and located at an equatorial spacing of 38 nm (position Q in Fig. 1a) where the streak and disk are recorded together. Resting profile (lower curve) and change in profile (upper curve) at peak of tension in a phasic ABRM contraction at 28 °C in sea water with serotonin. Note that the disk intensity (between single and double arrows) has decreased and that the streak intensity (inside double arrows) after subtraction of the disk-background has increased. The justification for dividing the data into two components is that during contraction the intensity at any point outside the double arrows follows one particular time course (equal to that shown in Fig. 3a, $\diamond$) and at any point inside it follows a completely different time course. (The main features of the latter are a sharp peak that occurs about 2 s after the beginning of stimulation, followed by a pronounced minimum which practically coincides with peak tension and after that there occurs a second broader peak.) After subtraction of the background formed by the disk, the time course is equal to that shown in Fig. 3a, Δ . The disk background was estimated as the intensity in 10 channels just outside the double arrows on both sides of the profile. *b*, Intensity profiles recorded with the detector along the equator (position P in Fig. 1a) and rest (R) and at peak of tension in phasic contractions (C). These data were accumulated from experiments on four different muscles. The 14-nm peak appears on a fairly steep background, coming mainly from the streak and the disk. The time course of the integrated intensity of the peak after background subtraction during phasic contraction is shown in Fig. 4b. The background was defined as the best second-order polynomial fitted to the intensity on either side of the peak. Further analysis of the peak profile shows that as its intensity decreases, no changes in its equatorial width take place but that its spacing can decrease up to 3%. The latter indicates that the actin rosette moves closer to the thick filament surface when cross-bridges are formed during generation of tension.

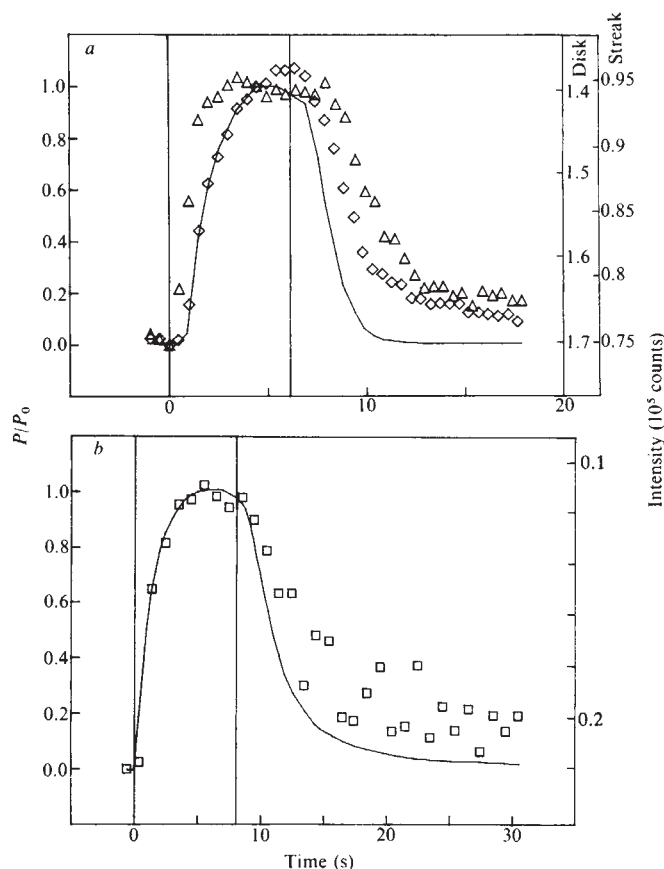


Fig. 3 *a*, Time courses of tension (—), streak intensity (Δ) and disk intensity ($\diamond$) recorded with the detector as shown in Fig. 1a at position Q during a phasic contraction at 28 °C in seawater containing serotonin. The period of stimulation is indicated by two vertical lines. The time resolution was 0.5 s. Streak intensity was integrated between the double arrows shown in Fig. 2a after subtraction of the background defined by the disk (outside double arrows). Disk intensity was integrated between the single and double arrows as shown in Fig. 2a. Note that the streak intensity increases much faster than tension and has reached a plateau within 2 s of the start of stimulation. Streak intensity remains at this plateau value until tension has decayed to about 20%. When tension is zero the streak intensity is only halfway back to its resting value. Disk intensity also takes longer than tension to return to its resting value but during the rising phase disk intensity decreases with a time course close to that of tension. *b*, Time course of tension (—) and integrated intensity of the 14-nm equatorial peak after background subtraction ($\square$) during a phasic contraction in natural seawater at 14 °C. The period of stimulation is indicated by two vertical lines. The time resolution was 1 s. Note that during the rising phase the intensity decreases with the same time course as tension increases whereas during relaxation the intensity recovers much more slowly than tension decays.

not due to a decrease in the packing order of the actin filaments. This is because analytical arguments¹⁴ show that in small one-dimensional lattices a decrease in intensity caused by disorder also leads to a widening of the peak in the equatorial direction. We have confirmed this argument by numerical modelling. Thus, it may be concluded that the observed intensity decrease is produced entirely by the extra mass added to the actin filaments as myosin heads.

The extent of the intensity decrease in the 14-nm peak depends on the peak tension. The maximal observed decrease was by 50%. To explain such a large change with the available myosin head mass relative to the actin filament mass (which as shown above is 0.29:1), the myosin heads must move to their most effective position, namely fully into the space between the actin filaments as shown in Fig. 1b at *f*. In that location we calculate that a mass ratio of 0.29 will cause a 50% decrease in the intensity of the 14-nm peak. This interpretation of the movement of myosin heads into the rosette spaces is supported by the data on the streak intensity changes.

The circular symmetry of the disk (Fig. 1a) indicates that the scatter is due to completely disordered units. The Guinier plot¹⁴

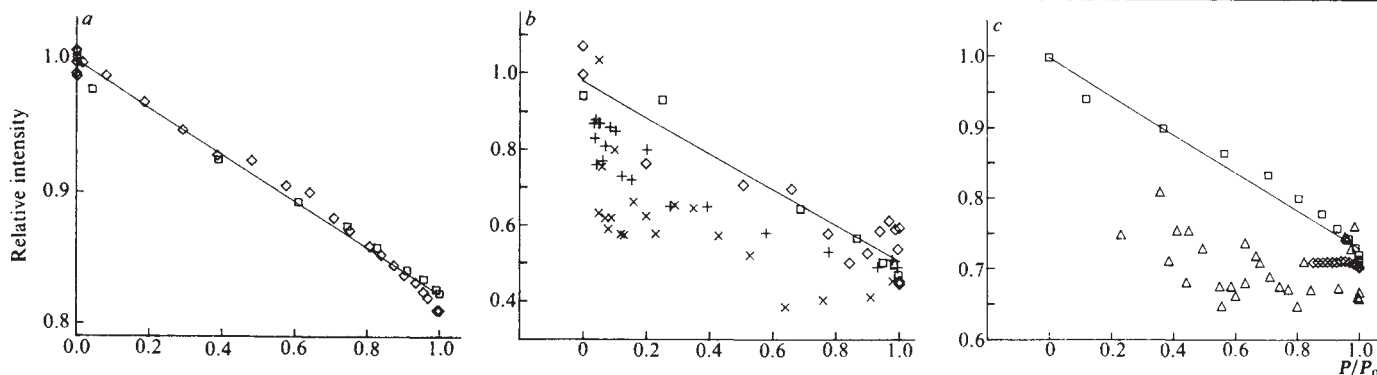


Fig. 4 *a*, Relationship between disk intensity and tension during the rising phase. Data taken from Fig. 3*a* ($\square$) and from an experiment carried out at 10 °C where the time course was about a factor of two slower ($\diamond$). Here and in *b* and *c* the regression line of intensity of tension is shown during the rising phase. *b*, Relationship between intensity of the 14-nm equatorial peak and tension during phasic contractions. Data for tension rise and relaxation phase ($\square$, +), respectively, were taken from Fig. 3*b* and from one other similar experiment ($\diamond$, $\times$) where the time resolution was 0.5 s and the time course of tension rise was identical to that shown in Fig. 3*b*. *c*, Relationship between intensity of the 14-nm equatorial peak and tension during tonic contraction produced by application of acetylcholine (5.5×10^{-6} M). In one experiment the time resolution was 2 s during both the rising phase ($\square$) and the catch phase of relaxation ($\diamond$) which follows washing out of acetylcholine. Further data from the relaxation phase of three other muscles are shown ($\triangle$) where the time resolution was 3 min. Except for the difference in time scale, the relationship between intensity and tension closely resembles that observed in phasic contractions illustrated in *b*.

gives a straight line (Fig. 5), indicating that the disk is due to one kind of scattering unit. The zero-angle intensity can be estimated by extrapolation of the line to its intercept with the intensity axis and is a measure of the number of scattering units¹⁴. In the experiment illustrated (Fig. 5), the observed change in the zero-angle intensity corresponds to a 35% decrease in that number at peak tension. The slope of the line is directly related to the radius of gyration R_g (ref. 14). As the slope does not change during contractions it follows that this also applies to the size of the scattering unit. Moreover, it means that the integrated intensity of any section of the disk gives a direct measure of the number of scattering units.

The close similarity between the time course of tension and change of intensity of the disk during the rising phase of contraction (Fig. 3*a*) suggests that the scattering units are composed of disordered myosin heads that become ordered on attachment to actin sites. The R_g values determined from the slope of the Guinier plots ranged from 8 to 16 nm. These values are similar to those obtained for heavy meromyosin from light scattering studies (14.7–18.5 nm, ref. 22) and suggest that in the ABRM the scattering unit is a dimer of myosin heads.

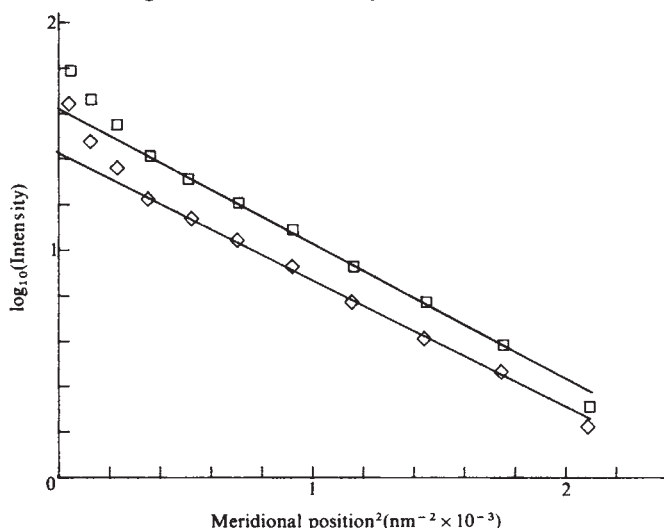


Fig. 5 Guinier plots¹⁴ of intensity profiles recorded with detector positioned as shown in Fig. 1*a* at Q; before stimulation ($\square$) and at peak tension ($\diamond$) in a phasic contraction at 14 °C in the presence of serotonin. The linear parts of the plots correspond to the diffuse disk. See text for summary of evidence on which it is suggested that the disk is due to disordered myosin heads. The deviation from the linear relation at small values of the square of the meridional position is caused by the superposition of the streak on the disk (see Fig. 2*a*). The slope of the straight line we have determined from the off-centre section through the disk is the same as that which would be derived from a section through the centre. Since the slope does not change during contraction, the intercept of the line with the intensity axis is directly proportional to the zero-angle intensity.

Discussion

Having reached *d* in Fig. 1*b*, the myosin heads could move into the rosette spaces in various different ways. Two extreme possibilities are that all heads move simultaneously towards their final position at *f* or that they move one-by-one fully into this position. Any significant contribution from the first mode of movement can be excluded because this would lead to a severe distortion of the shape of the 14-nm equatorial peak during tension rise. Our results during tension rise show no shape changes from those illustrated in Fig. 2*b*. Thus, we arrive at a structural mechanism for the movement of myosin heads in and out of the rosette spaces which can be represented by a two-state structural model where cycling myosin heads spend most of their time being either fully inside or fully outside the rosette spaces. Assuming that our identification of the disordered scattering units as myosin heads is correct, further support for the proposed model can be deduced from the finding that the number of disordered scattering units decreases during tension rise with the same time course as that of the increase in the number of myosin heads accumulating in the rosette spaces (Fig. 3*a,b*). This would indicate that during tension rise the myosin heads go directly from the disordered state into the ordered state in the rosettes.

The change in intensity of the equatorial streak has a significant lead over tension (Fig. 3*a*). In muscles like the ABRM it is known that the myosin heads are switched on by an action of Ca^{2+} on the light chains²³. We suggest that in this system, Ca^{2+} release also has a direct effect on the transfer of myosin heads from the thick filament surface to actin and that this effect is active and unidirectional. In other words, Ca^{2+} release does not bring about the transfer of myosin heads by shifting a passive equilibrium towards actin as an initial small fraction of myosin heads (already at actin) are switched on and attach. If that happened the rate of myosin head transfer would not exceed that of tension rise. It therefore appears that the time course of the streak intensity increase could give a direct measure for the spread of activation.

During the rise of tension it seems likely that the fall of intensity in both the 14-nm peak and in the disk are due to the close attachment of myosin heads to the actin filaments, and our model (Fig. 1*b*) suggests that the heads are then located between neighbouring filaments. The close similarity of time course of the tension and intensity changes implies that tension rises in direct proportion to the number of attached myosin heads.

Our results suggest an interesting structural explanation for the catch mechanism. During catch relaxation (Fig. 4*c*) the 14-nm peak intensity-tension relation is completely different from the linear one observed during tension rise. It appears that myosin heads remain in the rosette spaces for a

considerable time after they stop their chemical cycle, and this could explain how passive tension is held during catch^{8,9}. If, but for a difference in time scale, the same happens during phasic relaxation (Fig. 4b), this would also explain why here 'active state' decays faster than tension⁸. Serotonin would then act to speed up the release of passive myosin heads from the rosette spaces in both catch and phasic contractions.

On our interpretation of the disk scattering, the results presented in Figs 3a and 4a show that during tension rise the number of disordered myosin heads decreases with the same time course as the number of attached ones increases. From this it is conceivable that the disordered myosin heads could represent the detached phase of the cross-bridge cycle. If that can be confirmed, further analysis of our data will produce estimates for rate constants of attachment and detachment of myosin heads in the cross-bridge cycle.

Taking all the results and interpretations, the sequence of structural events during the mechanical contractile cycle may be summarized as follows (Fig. 1b, quadrants I–IV). On activation, disordered myosin heads move away from the paramyosin core surface towards the actin filaments (IV–I), causing an

increase in streak intensity but no change in disk intensity. As myosin heads start cycling they move into the spaces between the actin filaments (I–II), causing a decrease in intensity of both the 14-nm peak and the disk because such myosin heads are no longer disordered enough to contribute to the disk; a further increase in the streak intensity will also occur. When tension has decayed to a low level myosin heads leave the rosette spaces (II–III), causing the 14-nm peak and disk intensities to recover. Eventually, myosin heads move back to the core surface (III–IV) and streak intensity returns completely to its resting level.

We thank Drs J. Bordas and M. H. J. Koch (EMBL Outstation, Hamburg) for their most effective and enthusiastic collaboration in carrying out the experimental work and data reduction; Drs R. H. Abbott, H. E. Huxley and R. M. Simmons for many useful discussions; and Professor R. H. Stuhmann (Head of Outstation, Hamburg), Dr D. A. Blackburn (Director of Oxford Research Unit), Professor G. F. Elliott (Head of Biophysics Group, Oxford Research Unit) and other members of the staff in Hamburg and Oxford for their support. This work was supported by the MRC and the Muscular Dystrophy Association of the USA.

Received 27 October 1981; accepted 1 July 1982.

1. Huxley, H. E. *Science* **164**, 1356–1366 (1969).
2. Huxley, A. F. *Prog. Biophys. biophys. Chem.* **7**, 255–318 (1957).
3. Lowy, J. & Hanson, J. *Physiol. Rev.* **42**, 34–47 (1962).
4. Sobieszek, A. J. *Ultrastructure* **43**, 313–343 (1973).
5. Lowy, J. & Vibert, P. J. *Nature* **215**, 1254–1255 (1967).
6. Fletcher, C. M. J. *Physiol., Lond.* **91**, 172–185 (1937).
7. Twarog, B. M. J. *cell. comp. Physiol.* **44**, 141–163 (1954).
8. Lowy, J. & Millman, B. M. *Phil. Trans. R. Soc. B246*, 105–148 (1963).
9. Jewell, B. J. *Physiol., Lond.* **149**, 154–177 (1959).
10. Baguet, F. & Gillis, J. M. J. *Physiol., Lond.* **198**, 127–143 (1968).
11. Cornelius, F. & Lowy, J. *Acta physiol. scand.* **102**, 167–180 (1978).
12. Huxley, H. E., Faruqi, A. R., Milch, J. R., Bordas, J. & Koch, M. H. J. *Nature* **284**, 140–143 (1980).

13. Bordas, J., Koch, M. H. J., Clout, P. N., Boulin, C. & Gabriel, A. J. *Phys. scient. Instrum.* **E13**, 938–944 (1980).
14. Guinier, A. in *X-Ray Diffraction* 326–333 (Freeman, San Francisco and London, 1963).
15. Hanson, J. & Lowy, J. *Proc. Roy. Soc. B154*, 173–196 (1961).
16. Elliott, G. F. J. *molec. Biol.* **10**, 89–104 (1964).
17. Elliott, A., Lowy, J., Parry, D. A. D., & Vibert, P. J. *Nature* **218**, 656–659 (1968).
18. Elliott, A. & Lowy, J. J. *molec. Biol.* **53**, 181–203 (1970).
19. Elliott, A. J. *molec. Biol.* **132**, 323–341 (1979).
20. Elliott, A. *Proc. R. Soc. B186*, 53–66 (1974).
21. Mendelson, R. & Kretschmar, K. M. *Biochemistry* **19**, 4103–4108 (1980).
22. Holtzer, A., Lowy, S. & Schuster, T. in *The Molecular Basis of Neoplasia* 259–280 (University of Texas Press, Austin, 1962).
23. Kendrick-Jones, J., Lehman, W. & Szent-Györgyi, A. J. *molec. Biol.* **54**, 313–326 (1970).
24. Vainshtein, B. K. in *Diffraction of X-Rays by Chain Molecules* 125–143 (Elsevier, Amsterdam, 1966).

Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions

Charles K. Singleton*, Jan Klysik*, Steven M. Stirdivant* & Robert D. Wells*

Department of Biochemistry, College of Agricultural and Life Sciences, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA

In physiological ionic conditions (200 mM NaCl), the (dC–dG)₁₆ and (dC–dG)₁₃ blocks in plasmid pRW751 are in a left-handed state when the negative superhelical density of the plasmid is greater than 0.072. As the salt concentration decreases or when (d^mC–dG) sequences are present, less negative supercoiling is required to induce the right- to left-handed DNA transition. Furthermore, the single strand-specific nuclease, S₁, recognizes and cleaves aberrant structural features at the junction between neighbouring right- and left-handed DNA regions.

RECENT studies^{1–3} have demonstrated that alternating (dC–dG) sequences can adopt a left-handed helical conformation in negatively supercoiled plasmids in high salt (several molar) conditions. Furthermore, the torsional strain of superhelical DNA can enhance the salt-induced transition from a right-handed to a left-handed helical state². This is consistent with the unfavourable free energy of superhelix formation^{4,5}. Thus, processes which bring about relaxation of negative superhelical density such as a right- to left-handed helical transition have a favourable free energy input from the reduction of this torsional strain.

Alternating (dC–dG) sequences also are self-complementary (that is they have inverted repeat symmetry) and thus in principle are able to form hairpin loops or cruciforms. It has been shown recently that supercoiling can stabilize and bring about the formation of cruciforms at inverted repeats of several circular DNAs^{6–8}. Thus, it was of interest to determine if the 32-

and 26-base pair (bp) (dC–dG) stretches of pRW751 (ref. 1) would adopt a cruciform structure in low salt conditions. This was tested by using single strand-specific nucleases (S₁ and gene 3 product of phage T₇) as probes for the cruciform state^{6–8}. Surprisingly, we found that cruciforms are not formed in the two (dC–dG) blocks but instead, a right- to left-handed primary helical transition is induced solely by the torsional strain of supercoiling at millimolar salt concentrations. This transition is sensitive to the salt concentration and C5 methylation within the (dC–dG) blocks such that decreasing salt concentrations and methylation allow the transition to occur at lower negative superhelical densities. Also, we show that S₁ nuclease recognizes structural features of the junction between adjacent left-handed and right-handed helical segments.

S₁ nuclease treatment of pRW751

The plasmid pRW751 has been characterized previously^{1,9}. Briefly, this plasmid is a pBR322 derivative containing an insert of 157 base pairs into the *Bam*HI site. The insert consists of a 95-bp *Escherichia coli* lac operator DNA fragment flanked by

* Present addresses: University of Alabama, Birmingham, Schools of Medicine and Dentistry, Department of Biochemistry, Birmingham, Alabama 35294, USA (C.K.S., J.K., R.D.W.); The Biological Laboratories, Harvard University, Cambridge, Massachusetts 02138, USA (S.M.S.).

alternating (dC-dG) stretches of 32 and 26 bp (Fig. 2). Samples of this plasmid with various negative superhelical densities were generated and characterized as described previously¹⁰. Agarose gel electrophoresis of several of these samples is shown in Fig. 1a, and the calculated mean negative superhelical density of each is listed in Fig. 1 legend. As the negative density increases, the gel mobility increases.

Each sample was treated with *S*₁ nuclease, followed by *Ava*II restriction digestion, to determine whether supercoiling could induce any single strand structural feature(s) which *S*₁ nuclease might recognize and cleave. Figure 1b illustrates that several

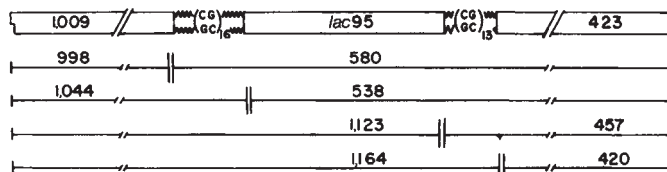


Fig. 2 Fragment sizes generated by *S*₁ nuclease-specific cleavages near the (dC-dG) blocks of pRW751. Sample 9 of Fig. 1a, b was digested first with *S*₁ nuclease and then with *Ava*II as described in Fig. 1 legend. The resulting fragments were separated by 2% agarose gel electrophoresis together with the appropriate size markers. The top line represents the (dC-dG)-containing 1,585-bp *Ava*II restriction fragment. Within this fragment, the sizes of non-(dC-dG) as well as the (dC-dG) sequences are indicated. The sizes of the *S*₁-specific bands derived from the *Ava*II 1,585-bp band were determined²⁹ and are illustrated by the bottom four lines. The larger bands are correct to within ± 10 bp, while the accuracy for the smaller bands is ~ 7 bp. The known length of fragments possessing an *Ava*II terminus and ending at a non-(dC-dG)/(dC-dG) interface (junction) are (from left to right): 1,009 + 576; 1,041 + 544; 1,136 + 449; 1,162 + 423. Mapping of the bands arising from *S*₁ cleavage at the major pBR322 cruciform gave 1,300 bp (1,306 expected) and 437 bp (436 expected).

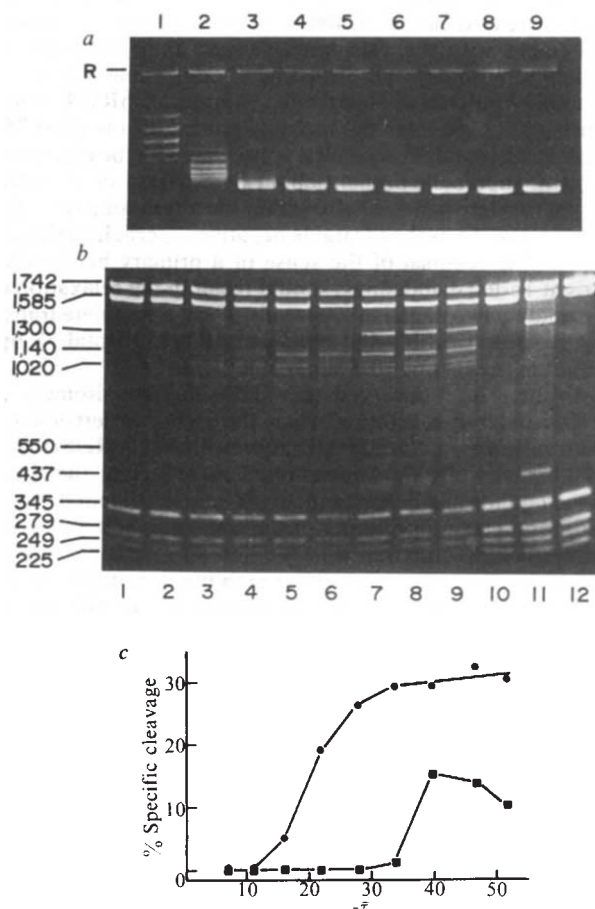


Fig. 1 Cleavage of pRW751 with *S*₁ nuclease followed by *Ava*II digestion. **a**, 1% agarose gel electrophoresis (80 mM Tris-acetate, 40 mM Na acetate, 2.5 mM EDTA pH 8.3, 85 V, 18 h) of various supercoiled samples of pRW751. The samples were generated and characterized as described elsewhere¹⁰. Plasmid DNA was isolated as described previously⁹. The mean negative titratable superhelical density in *S*₁ digestion conditions was: 1, 0.015; 2, 0.024; 3, 0.035; 4, 0.048; 5, 0.062; 6, 0.075; 7, 0.088; 8, 0.104; 9, 0.115 (calculated using 10 bp per turn and pRW751 = 4,518 bp). R, relaxed plasmid. **b**, Aliquots (1.5 μ g) of each sample shown in **a** were treated with 0.08 units of *S*₁ nuclease²⁸ in 50 μ l of 40 mM Na acetate, 50 mM NaCl, 1 mM ZnSO₄, pH 4.6. Reactions were carried out at 37°C for 50 min then stopped by addition of 2 μ l of 500 mM EDTA-2Na, pH 8. Following dialysis against 15 mM Tris-HCl, 6 mM NaCl, 6 mM MgCl₂, 2.5 mM dithiothreitol, pH 7.7, one unit of *Ava*II (BRL) was added and digestion was carried out at 37°C for 5 h. The DNA fragments were then separated by 2% agarose gel electrophoresis (same buffer as above, 175 V, 5 h). Lanes 1-9 correspond to the supercoiled pRW751 samples shown in **a**; lane 10 is pRW751 without *S*₁ treatment; lanes 11 and 12 are control plasmid pRW451 (ref. 1) treated with and without *S*₁, respectively. 1,140, 1,020 and 550 bp represent the approximate centre for each doublet. **c**, Specific cleavage by *S*₁ nuclease plotted against $-\tau$ (mean number of negative superhelical turns). A photographic negative of **b** was traced using a Joyce-Loebl microdensitometer, and the area of each peak was determined in conditions of linearity. Specific cleavage near the (dC-dG) blocks was obtained from the ratio of the 1,585-bp band (containing the (dC-dG) inserts) to the 1,742-bp band relative to this ratio for the sample not treated with *S*₁ (●). Specific cleavage at the major pBR322 cruciform site^{6,7} was determined analogously after *Taq*I digestions and analysis of loss of the fragment containing this site (■).

*S*₁-specific bands appear as the negative density increases. These bands were not present in the controls which were not treated with *S*₁ nuclease (Fig. 1b, lanes 10, 12).

The first bands which appeared were four sets of doublet bands centred around 1,140, 1,020, 550 and 430 base pairs. Concomitant with the appearance of these bands, the *Ava*II 1,585-bp band decreased in intensity compared with the other *Ava*II bands. The 1,585-bp band contains the (dC-dG)-flanked insert of pRW751. Also, as the doublet bands were not seen when pRW451 (ref. 1; control plasmid of about the same size as pRW751 but not containing any alternating (dC-dG) sequences) was treated with *S*₁ nuclease, these bands are specific for the (dC-dG)-containing insert. At higher negative superhelical densities, two new bands appeared; these can be assigned to *S*₁ cleavage at the major pBR322 cruciform site^{6,7} based on their mobilities and their production when the control plasmid, pRW451, was treated with *S*₁ nuclease (Fig. 1b, lane 11).

The per cent specific cleavage (relative loss of the restriction fragment containing the 157-bp insert or the pBR322 cruciform site) by *S*₁ nuclease at these sites was plotted against the number of mean negative superhelical turns for each sample (Fig. 1c). Specific cleavage within the 1,585-bp band reached a maximum at a negative superhelical density of ~ 0.07 and remained at this level as the negative density increased.

Mapping of *S*₁ nuclease-specific cleavage sites

The *S*₁ nuclease-specific fragments shown in Fig. 1b were sized by gel electrophoresis of an *S*₁ nuclease-treated, *Ava*II-digested sample, together with markers of the appropriate lengths. The results of this mapping are illustrated in Fig. 2. For the eight bands derived from the *Ava*II 1,585-bp fragment, *S*₁ nuclease cleavage occurred in the regions flanking the (dC-dG) blocks, within 2-13 bp of the first alternating (dC-dG) residues. This pattern of cleavage explains the appearance of doublet bands; the larger band of each doublet represents cleavage near the boundary of a (dC-dG) block distal to an *Ava*II site while the smaller band resulted from cleavage near the proximal boundary. The average length difference for the two bands of each doublet was 44 bp for the doublets corresponding to cleavage around the 32-bp (dC-dG) block and 39 bp for those flanking the 26-bp (dC-dG) block. Mapping of the *S*₁ nuclease cleavage sites with other restriction endonucleases (*Pst*I, *Eco*RI and *Alu*I) was in complete agreement with the *Ava*II mapping data.

Models consistent with doublet pattern

At least two models are consistent with this pattern of cleavage. The first would be cleavage by *S*₁ nuclease at the junction between the normal helix and the stem of a cruciform. This model seems unlikely, however, as *S*₁ nuclease cleavage at cruciforms has, so far, always mapped as a single band consistent

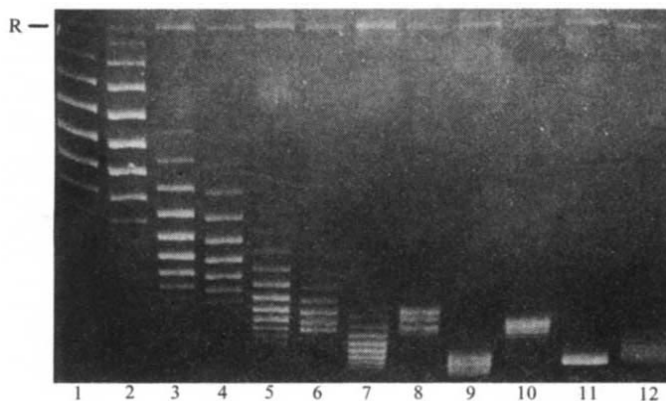


Fig. 3 Gel electrophoretic analysis of supercoiled samples of pRW751 (even-numbered lanes) and pRW451 (odd-numbered lanes). 2% Agarose gel electrophoresis was carried out using the buffer system described in Fig. 1 legend. Electrophoresis was at 140 V for 29 h. The calculated mean number of negative supercoils (and that determined by gel electrophoresis) is: lane 1, 5.3(5); 2, 5.3(5); 3, 9.1(9); 4, 10(10); 5, 13.3(13.5); 6, 14.6(14); 7, 17(18); 8, 18.1(15); 9, 22(21.5); 10, 23.9(16); 11, 26.5(-); 12, 29.4(18.5). A value for sample 11 was not obtained due to the lack of resolution of topoisomers at this high density. R, relaxed plasmid.

with cleavage within the single strand loop region (Fig. 1b)^{6,7}. Even though alternating (dC-dG) sequences are not interrupted with a region of non-inverted symmetry, steric constraints argue that a loop of at least four bases must occur^{11,12}. A cruciform within the (dC-dG)₂₆ block would thus have a loop of 4 bases and a stem of 11 bp; such a cruciform should be comparable in stability to the pBR322 major cruciform. However, the results presented in Fig. 1 show that the S₁ nuclease-specific bands corresponding to cleavage at the 26-bp (dC-dG) boundaries appear at much lower negative superhelical densities than the bands corresponding to cleavage at the pBR322 cruciform site.

A second model consistent with the cleavage pattern observed is that the torsional strain of supercoiling induced a right- to left-handed helical transition of the (dC-dG) blocks. Such a transition would partially relieve this torsional strain. As the unfavourable free energy of superhelix formation increases quadratically with increasing negative superhelical density^{4,5}, this transition would be more favourable at higher negative densities. Because only the (dC-dG) sequences in the plasmid can undergo this transition^{1,2,13}, a junction must be present between the left-handed helical (dC-dG) sequences and the right-handed helical flanking sequences¹. If this junction possesses any single-strandedness, S₁ nuclease may be recognizing and cleaving at these sites. Although little is known about the structural details of a right-handed-left-handed helical junction, model building studies have suggested that at least one base pair 'slip-out' is required within the junction due to steric constraints¹⁴. As discussed below, other data support this second model.

Gel analysis reveals relaxation of negative supercoils

pRW751 samples possessing a gaussian distribution of about eight topoisomers were used to obtain the results presented in Fig. 1. These samples were produced and characterized by a method which, in essence, determines the average amount of unwinding of the two DNA strands relative to the normal amount of winding for the relaxed plasmid¹⁰. The unwinding is brought about by intercalation of ethidium molecules into the DNA duplex in the presence of a topoisomerase. By determining the amount of bound ethidium, the level of unwinding can be calculated by assuming each ethidium unwinds the primary helix by 26° (ref. 15). This unwinding represents the mean linking difference¹⁶ for the plasmid sample and is partitioned between twist, writhe and other structural features in a

manner that minimizes the free energy of the unwound state¹⁷⁻¹⁹ once the ethidium is removed. The mean superhelical density (or number of superhelical turns) of each sample is calculated assuming that the DNA is in a normal B helix and all unwinding is partitioned into negative superhelical turns¹⁰. Thus, these values are the classical titratable superhelical densities²⁰.

When samples of pRW751 having different superhelical densities were analysed by 2% agarose gel electrophoresis (Fig. 3, even-numbered lanes), a discrepancy arose for the samples that were more highly supercoiled; gel analysis gave a lower estimate of the mean negative superhelical density than the calculated values. Such a discrepancy was not seen for the control plasmid, pRW451, for which the mobilities of the samples increased as the titratable negative density increased, in agreement with the calculated values, until resolution of individual topoisomers was lost (Fig. 3, odd-numbered lanes). The mobility pattern of supercoiled samples of pRW451 was as expected^{10,21}, whereas the mobility pattern of the pRW751 samples was atypical. This result is consistent with the relaxation of negative supercoils brought about by a partial or complete transition to a left-handed helix within the alternating (dC-dG) blocks of pRW751 as the titratable negative superhelical density increases. The reversal of the sense of a primary helical turn from right-to-left-handedness should result in the relaxation of approximately two negative supercoils. Thus, a complete transition within the two (dC-dG) blocks of pRW751 (total 58 bp) will result in the loss of 10.5 supercoils².

Stirdivant *et al.*² observed that individual topoisomers of pRW751 undergo relaxation when the right- to left-handed helical transition is induced by titration with NaCl. Relaxations of <10.5 supercoils were interpreted as reflecting the time-averaged equilibrium between the right- and left-handed states. Indeed, for a given topoisomer, if the transition between these two states is fast compared with the time of electrophoresis, a sharp band will be observed with a mobility dependent on the time-averaging of the superhelical density of the two states⁵. As the samples used here consisted of a gaussian distribution of about eight topoisomers, following the relaxation behaviour was slightly more complicated. Within a given sample, the topoisomers which are more negatively supercoiled will relax to a greater extent than those having less supercoiling. Thus, the expected pattern would be a compaction and apparent smearing of the topoisomer population as the equilibrium shifts in favour of the left-handed helical state (that is, as the negative superhelical density increases). Finally, the gaussian distribution

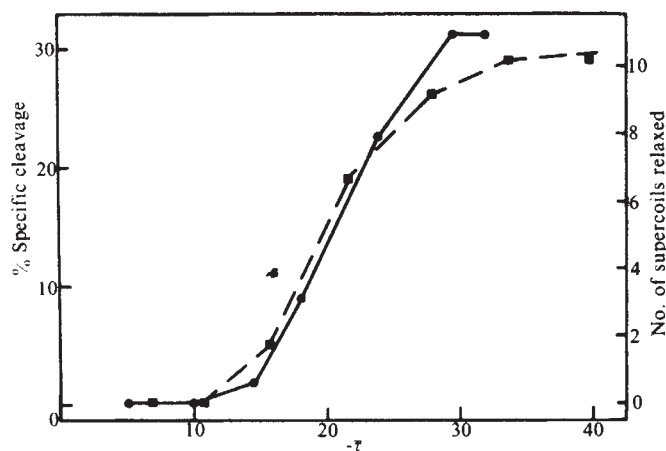


Fig. 4 Average number of negative supercoils relaxed plotted against the mean number of titratable superhelical turns, $-\bar{\tau}$. The number of supercoils relaxed for pRW751 was obtained by determining the median position of the topoisomer distribution from Fig. 3, and subtracting this value from the expected median position based on the calculated titratable superhelical density (●). The transition end points are based on analysis of the behaviour of individual topoisomers within a given sample. The broken line (■) illustrates S₁ nuclease-specific cleavage versus $-\bar{\tau}$ (see Fig. 1c).

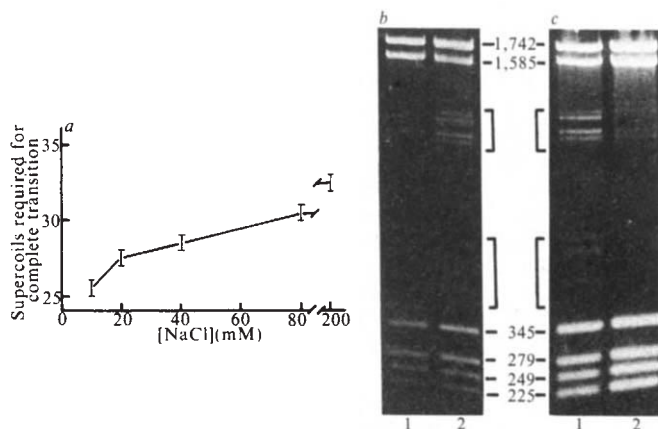


Fig. 5 Effect of Na^+ and methylation on formation of left-handed DNA in pRW751. *a*, Mean number of titratable supercoils required to give the complete right- to left-handed transition plotted against NaCl concentration of the electrophoresis buffer. The electrophoresis buffer (40 mM Tris-HCl pH 8.3, 1.25 mM EDTA, 10–200 mM NaCl) was continuously circulated. Electrophoresis of supercoiled pRW751 samples was carried out on a 2% agarose slab gel for 27–52 h. The current was regulated such that the temperature of each gel was $\sim 28^\circ\text{C}$. The transition end point was determined as described in Fig. 4 legend. *b*, *Ava*II digest of *S*₁ nuclease-cleaved pRW751. *S*₁ nuclease digestion conditions were those outlined in Fig 1 legend (lane 1); lane 2, the sample was digested in low salt conditions: 40 mM Na acetate, 1 mM ZnSO_4 , pH 4.6 *c*, *Ava*II digest of *S*₂-cleaved methylated (lane 1) and unmethylated (lane 2) pRW751. Methylation was carried out essentially as described elsewhere³⁰ with *Hha*I methylase (G^mC–G–C). Methylation was monitored by observing loss of cleavage with *Hha*I restriction endonuclease (J. K., unpublished). Brackets denote *S*₁ nuclease-specific bands.

should reappear at a position indicative of a loss of ~ 11 negative supercoils once the mean superhelical density is sufficient to cause the complete transition.

Compaction and smearing of the topoisomer population was found for the pRW751 samples 6–10 of Fig. 3. The population of topoisomers shown in Fig. 3, lane 12 appears as a normal distribution of topoisomers; however, the average number of negative supercoils is about 11 less than the expected, calculated value. The average number of supercoils for the first two pRW751 samples shown in Fig. 3 (and all but the two highest supercoiled topoisomers of sample 3) were as expected, indicating that these topoisomers are of insufficient superhelical density to induce the right- to left-handed helical transition.

Figure 4 illustrates the supercoil-induced relaxation behaviour of pRW751. Relaxation of pRW751 plateaus after a loss of about 11 negative supercoils. This degree of relaxation correlates well with the expected value for a complete transition from a right- to a left-handed helix within the (dC–dG) blocks, but is inconsistent with cruciform formation at these sites, a process which would result in the relaxation of only about six supercoils as 58 bp would be unwound on cruciform formation.

These results demonstrate that at or above a titratable negative superhelical density of 0.064, the (dC–dG) blocks within pRW751 are completely in a left-handed helical state in the gel buffer conditions. This result is also consistent with the levelling off of *S*₁ nuclease-specific cleavage around a density of ~ -0.07 . Indeed, as shown in Fig. 4, specific cleavage by *S*₁ nuclease correlates well with relaxation as revealed by gel analysis.

When isolated from *E. coli* cells, pRW751 possesses a lower negative superhelical density than the control plasmid, pRW451, and other pBR322 derivatives as revealed by gel electrophoretic analysis². By direct comparison with the gel results presented here (Fig. 3), we have determined that part of the reduced density can be attributed to a relaxation of supercoils due to formation of left-handed DNA within the (dC–dG) blocks (data not shown). This is also consistent with

pRW751 as isolated from *E. coli* cells showing *S*₁ nuclease-sensitivity at the (dC–dG) blocks.

Salt effects

When the supercoiled samples of pRW751 were electrophoresed in different ionic conditions, the number of negative supercoils required to induce complete relaxation decreased as the NaCl concentration decreased (Fig. 5*a*). This effect is greatest at low salt concentrations and appears to level off in the 200 mM range. Figure 5*b* illustrates that specific cleavage by *S*₁ nuclease is also enhanced at a lower Na^+ concentration. The enhancement of specific cleavage on decreasing the Na^+ level was observed only for samples which had not undergone the complete transition. Thus, this effect results in a shift to the left of the curves in Fig. 4.

This salt effect was unexpected as Stirdivant *et al.*² found the opposite relationship between salt concentration and superhelical density for the formation of left-handed DNA in pRW751. However, this previous study focused on the salt-induced transition from right- to left-handed DNA and thus used much higher concentrations (several molar) of NaCl. The results presented in this paper focus on the right- to left-handed DNA transition which is driven by the relief of the torsional strain of supercoiling at relatively low levels of sodium ions. Thus, the ionic environment of the DNA in the two studies was very different, which could account for the apparent discrepancy. Furthermore, the salt effect described here levelled off at higher Na^+ concentrations, suggesting a loss in the trend observed at lower salt concentrations.

Effect of methylation

It was demonstrated recently that C5-methylation of cytidine residues in (dC–dG)_n · (dC–dG)_n greatly facilitated the formation of a left-handed structure for this polymer^{22,23}. We have found that, although not as dramatic as with the polymer, methylation of the cytidine residues of the (dC–dG) blocks of pRW751 by *Hha*I methylase facilitates formation of left-handed DNA in the plasmid. Figure 5*b* illustrates that *S*₁ nuclease-specific cleavage is enhanced when the fully methylated plasmid is used as template. Gel analysis of the methylated plasmid DNA also demonstrated that less supercoiling is required to bring about formation of left-handed DNA relative to the unmethylated plasmid (data not shown). In general, the effect of methylation on left-handed DNA formation in pRW751 is equivalent to increasing the number of negative supercoils by four to five.

Conclusions

We have demonstrated that left-handed DNA can exist in plasmids in physiological conditions of ionic strength and superhelical density. Furthermore, we have shown that *S*₁ nuclease can serve as a probe for the conformational junction which must exist between adjacent left-handed and right-handed helical DNA regions. Specific cleavage within the junction region suggests that some degree of single strandedness exists within this region, which is consistent with model building studies¹⁴. It is apparent, though, that *S*₁ nuclease binds to and cleaves within the junction region rather inefficiently. This is reflected in the fact that specific cleavage by *S*₁ nuclease levels off at $\sim 30\%$ even though gel analysis of the plasmid samples demonstrates that the (dC–dG) blocks are completely stabilized in a left-handed helical state. That a relatively large region of single strandedness is not present is also consistent with the fact that we have been unable to demonstrate junction recognition by another single strand-specific nuclease, the gene 3 product of phage T₇ (data not shown).

The *S*₁ nuclease-specific fragments which correspond to cleavage near the 32-bp (dC–dG) block appear at lower negative superhelical densities than the bands corresponding to cleavage at the 26-bp block (Fig. 1*b*, lanes 3, 4). This is consistent with the previous finding² that salt-induced left-handed DNA formation within the 32-bp (dC–dG) region occurs at

lower salt concentrations than left-handed DNA formation within the 26-bp region. Note also that one junction (third from the left in Fig. 2) is recognized and cleaved by S_1 nuclease more readily than the other three, as judged by the band intensities in Fig. 1b. This result suggests significant differences between various junction regions, perhaps due to sequence effects. Finally, we wish to emphasize that S_1 nuclease cleavage apparently occurs outside the (dC-dG) blocks, implying that the complete (dC-dG) sequences are in a left-handed state, in agreement with previous studies^{1,2,24}.

The S_1 nuclease and gel analyses on samples of pRW751 containing different superhelical densities demonstrate that formation of left-handed DNA within the (dC-dG) blocks of this

plasmid begins at a negative superhelical density of ~ 0.039 and is completely stabilized at a density of -0.072 (18–32/33 negative supercoils) in physiological ionic conditions (200 mM NaCl). The range of superhelicity required to stabilize left-handed DNA in these conditions is well within the range of physiological superhelical densities typically found for most plasmids²⁰. If *in vitro* supercoiling is a true reflection of that occurring *in vivo*^{25–27}, our results indicate that left-handed DNA within (dC-dG) regions exists *in vivo* under the torsional strain of supercoiling.

This work was supported by NIH grant CA 20279 and NSF grant PCM-8002622. C.K.S. was supported by a postdoctoral fellowship (PF-1904) from the American Cancer Society.

Received 23 March; accepted 7 June 1982.

1. Klysik, J., Stirdivant, S. M., Larson, J. E., Hart, P. A. & Wells, R. D. *Nature* **290**, 672–677 (1981).
2. Stirdivant, S. M., Klysik, J. & Wells, R. D. *J. biol. Chem.* (in the press).
3. Zacharias, W., Larson, J. E., Klysik, J., Stirdivant, S. M. & Wells, R. D. *J. biol. Chem.* **257**, 2775–2782 (1982).
4. Pulleyblank, D. E., Shure, M., Tang, D., Vinograd, J. & Vosberg, H.-P. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4280–4284 (1975).
5. Depew, R. E. & Wang, J. C. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4275–4279 (1975).
6. Panayotatos, N. & Wells, R. D. *Nature* **289**, 446–470 (1981).
7. Lilley, D. M. J. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6468–6472 (1980).
8. Singleton, C. K. & Wells, R. D. *J. biol. Chem.* **257**, 6292–6295 (1982).
9. Klysik, J., Stirdivant, S. M. & Wells, R. D. *J. biol. Chem.* (in the press).
10. Singleton, C. K. & Wells, R. D. *Analyt. Biochem.* **122**, 253–257 (1982).
11. Tinoco, I., Uhlenbeck, O. C. & Levine, M. D. *Nature* **230**, 362–367 (1971).
12. Haasnoot, C. A. G., den Hartog, J. H. J., de Rooij, J. F. M., van Boom, J. H. & Altona, C. *Nucleic Acids Res.* **8**, 169–181 (1980).
13. Wells, R. D. *et al. J. biol. Chem.* (in the press).
14. Drew, H. R. & Dickerson, R. E. *J. molec. Biol.* **152**, 723–736 (1981).

15. Wang, J. C. *J. molec. Biol.* **89**, 783–801 (1974).
16. Wang, J. C. *Trends biochem. Sci.* **5**, 219–221 (1980).
17. Benham, C. J. *J. molec. Biol.* **150**, 43–68 (1981).
18. Camerini-Otero, R. D. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **75**, 1708–1712 (1978).
19. Vologodski, A. V., Anshelevich, V. V., Lukashin, A. V. & Frank-Kamenetskii, M. D. *Nature* **280**, 294–298 (1979).
20. Bauer, W. R. A. *Rev. Biophys. Bioengng* **7**, 287–313 (1978).
21. Keller, W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 4876–4880 (1975).
22. Behe, M., & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1619–1623 (1981).
23. Behe, M., Zimmerman, S. & Felsenfeld, G. *Nature* **293**, 233–235 (1981).
24. Wartell, R. M., Klysik, J., Hillen, W. & Wells, R. D. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2549–2553 (1982).
25. Gellert, M. A. *Rev. Biochem.* **50**, 879–910 (1981).
26. Pettijohn, D. E. & Pflüger, O. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1331–1335 (1980).
27. Sinden, R. R., Carlson, J. O. & Pettijohn, D. E. *Cell* **21**, 773–783 (1980).
28. Dodgson, J. B. & Wells, R. D. *Biochemistry* **16**, 2374–2379 (1977).
29. Southern, E. M. *Analyt. Biochem.* **100**, 319–323 (1979).
30. Mann, M. B. & Smith, H. O. in *Transmethylation* (eds Usdin, E., Borchardt, R. T. & Creveling, C. R.) 483–492 (Elsevier, New York, 1979).

Viroid RNAs of cadang-cadang disease of coconuts

James Haseloff, Nizar A. Mohamed^{†*} & Robert H. Symons

Adelaide University Centre for Gene Technology, Department of Biochemistry, University of Adelaide, Adelaide, South Australia 5001

* FAO/UNDP Coconut Research Project, Philippine Coconut Authority, Albay Research Center, Guinobatan, Albay, Philippines 4908

Cadang-cadang is a serious lethal disease of coconut palms in the Philippines. Infectivity is associated with four viroid RNAs (ccRNAs) which are from 246 to 574 residues in size but do not vary in sequence complexity. They share sequence and structural homology with other viroids. Variant ccRNAs found in nine isolates sequenced may have arisen as by-products or intermediates of viroid replication.

CADANG-CADANG is a serious and economically important disease of coconut palms which was first reported in 1927 on San Miguel Island in the Philippines¹. By 1962, all but 100 of the 250,000 palms on this island had died from the disease². Cadang-cadang disease has spread rapidly and now, 56 years after its first known occurrence, it is widespread over the south-east part of Luzon and many neighbouring islands (Fig. 1). It is estimated that cadang-cadang is responsible for the death of 500,000 palms each year in the Philippines (B. Zelazny, personal communication). Recent work has indicated that tinangaja disease of coconuts on Guam, an island in the Pacific 1,500 miles east of the Philippines, has the same aetiology as cadang-cadang³. However, cadang-cadang disease has not been found in any other coconut growing area.

Four low-molecular weight RNA species (ccRNAs) are found associated with cadang-cadang disease⁴. Two of these RNAs, called ccRNA 1 fast and ccRNA 2 fast, of ~ 250 and 500 nucleotides⁵ respectively, are present in infected palms at early stages of the disease⁴. As infection progresses over a period of years, two additional RNAs, ccRNA 1 slow and ccRNA 2 slow (of ~ 300 and 600 nucleotides, respectively), characterized by their slower electrophoretic mobilities on polyacrylamide gels,

appear and then predominate⁵. Ribonuclease fingerprinting showed that the ccRNAs, although differing in size, contain common repeated sequences, suggesting that the ccRNA 2 fast and slow species may be dimers of the respective ccRNA 1 species⁵ and that the three larger RNAs are closely related to ccRNA 1 fast.

All four ccRNA species have been shown recently to be infectious⁶. They are single-strand covalently closed circular RNAs with high degrees of secondary structure^{7,8} and so possess properties similar to those of a small group of infectious plant pathogens, the viroids⁹. Thus, the ccRNAs can now be classed definitely as the coconut cadang-cadang viroid (CCCV) in view of the recent infectivity data. The other viroids characterized each consist of single infectious RNA species which do not seem to code for functional polypeptide products but are capable of autonomous replication^{9,10}.

We report here the primary sequences and secondary structures of the four RNAs of several CCCV isolates. The results are compared with the structures of other viroids and are discussed in terms of the possible origin of cadang-cadang disease and mechanisms for the replication of viroids.

RNA sequence and structure determination

The approaches we have used successfully for the determination of the sequences and structures of chrysanthemum stunt viroid

[†] Present address: National Health Institute, PO Box 50348, Porirua, New Zealand.

Table 1 Properties of RNAs of various ccRNA isolates

ccRNA isolate*	Relative proportions of sequence variants†		Total length of ccRNA (residues)	Length of sequence duplication (residues)
	C	CC		
ccRNA 1 fast				
Baao 54	1.0	0	246	
Tinambac	1.0	0	246	
Ligao 14B	0.8	0.2	246/247‡	
Ligao 620C	0.6	0.4	246/247	
Ligao 191D	0.4	0.6	246/247	
Ligao T1	0.2	0.8	246/247	
ccRNA 2 fast				
Baao 54	1.0	0	492	
Ligao T1	0.2	0.8	492–494§	
ccRNA 1 slow				
Baao 54	1.0	0	287	41
Ligao 14B	1.0	0	296	50
Ligao 620C	1.0	0	296	50
Ligao 191D	1.0	0	296	50
Ligao T1	1.0	0	301	55
Ligao 5	1.0	0	296	50
Guinayangan	1.0	0	296	50
San Miguel	1.0	0	296	50
San Nasciso	0	1.0	297	50
ccRNA 2 slow				
Baao 54	1.0	0	574	41

* ccRNA isolates were purified⁴ from nucleic acid extracts³³ of single, infected coconut palms.

† Relative proportions of sequence variants were determined by sequence analysis of RNase T₁-digested, 5'-³²P-radiolabelled full-length linear ccRNAs. If a ccRNA consisted of a mixture of variants, band doubling was observed on sequencing gels¹⁰ after ccRNA 1 fast residue 197. Relative proportions of the two sets of band doublets were taken as estimates of the molar proportions of the two variant ccRNA species.

‡ These ccRNA 1 fast species consisted of a mixture of two species, one of 246 and the other of 247 residues.

§ Ligao T₁ ccRNA 2 fast species contained sequence heterogeneity at the positions corresponding to ccRNA 1 fast residue 198. Due to limitations of sequencing technique, we did not determine whether dimeric ccRNA 2 fast consisted of a 492(2×246)-residue species together with a 494(2×247)-residue species or only a 493(246+247)-residue species.

(CSV)¹⁰, avocado sunblotch viroid (ASBV)¹¹, citrus exocortis viroid (CEV)¹², and the viroid-like RNAs (virusoids) of velvet tobacco mottle virus (VTMoV)¹³ and solanum nodiflorum mottle virus (SNMV)¹³ were applied to the four ccRNAs. Native circular ccRNAs were subjected to limited digestion either by ribonuclease T₁, which catalysed cleavage of ccRNA 1 species at single sites and ccRNA 2 species at either or both of two sites to produce specific full-length linear ccRNAs, or by ribonucleases A or U₂, which produced smaller overlapping RNA fragments. These linear RNA molecules were 5'- or 3'-radiolabelled and then purified by polyacrylamide gel electrophoresis^{10,14}. The sequences of these 5'- or 3'-labelled fragments were determined by partial enzymatic digestion¹⁰. The use of fragments labelled separately at both the 5'- and 3'-ends allowed the sequence determination of long RNA fragments up to 574 residues long and, with shorter fragments, resolved gel compression artefacts¹⁵ when the relevant nucleotide sequences were determined from both directions. The sequences of overlapping fragments were assembled to construct the complete primary structures of the circular RNA molecules.

Secondary structure models for the native ccRNAs were constructed using the base pairing matrix procedure of Tinoco *et al.*¹⁶, values for the thermodynamic stability of predicted RNA structures (G. Steger, H. J. Gross, J. W. Randles, H. L. Sänger and D. Riesner, personal communication), and experimental evidence for the location of ribonuclease-sensitive single-stranded regions on the native molecules. In addition, specific full-length linear ccRNAs, produced by ribonuclease T₁ cleavage, were either 5'- or 3'-labelled with ³²P and the

single-strand regions in the native structures located by the nuclease S₁ mapping procedure¹⁷. The sites of cleavage were determined by co-electrophoresis of the radiolabelled fragments of the nuclease S₁ digest with products of sequencing reactions using the partial enzymatic digestion procedure¹⁰.

ccRNAs differ in size but not sequence complexity

The sequences and predicted structures of the ccRNA 1 fast and slow forms isolated from a single infected coconut palm (isolate Baao 54) are shown in Fig. 1 together with the known structures of four viroids, potato spindle tuber viroid (PSTV)¹⁸, CSV¹⁰, CEV¹², and ASBV¹¹. The two native ccRNAs 1 possess extensive regions of intramolecular base pairing and can form rod-like native structures similar to other viroids. The ccRNA 1 fast and ccRNA 1 slow possess 246 and 287 residues, respectively, and have calculated thermodynamic stabilities of -320 and -360 kJ mol⁻¹, respectively. ccRNA 1 slow contains the entire sequence and structure of the smaller ccRNA 1 fast but differs by an additional duplicated sequence and structure of 41 residues (residues 103–143 in ccRNA 1 fast) which is added at the right-hand end of the native molecule between residues 123 and 124 of ccRNA 1 fast (Fig. 2). Thus, the rod-like, base-paired native structure is maintained in the larger molecule.

The nucleotide sequences of ccRNA 2 fast and ccRNA 2 slow, consisting of 492 and 574 residues, respectively, are perfect dimers of the respective ccRNA 1 forms. A schematic summary of the relationships between the primary structures of all four ccRNAs and their predicted secondary structures is given in Fig. 3. While each of the monomeric ccRNA 1 forms can base pair intramolecularly to form a single rod-like conformer, the ccRNA 2 forms, due to their dimeric nature, can form either of two rod-like conformers (A or B, Fig. 3) and a large number of intermediate cruciform-shaped structures, one of which is given in Fig. 3.

The ccRNA 1 fast and ccRNA 1 slow molecules each possess, in experimental conditions, a highly accessible site for cleavage by ribonuclease T₁ at the right-hand terminal hairpin loop of the predicted native structure (between residues 124 and 125 in Baao 54 ccRNA 1 fast and between residues 145 and 146 in Baao 54 ccRNA 1 slow). Limited ribonuclease T₁ digestion

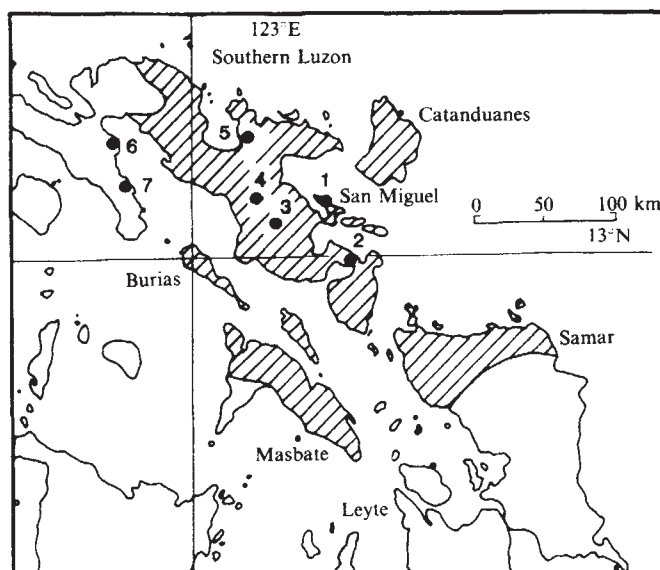
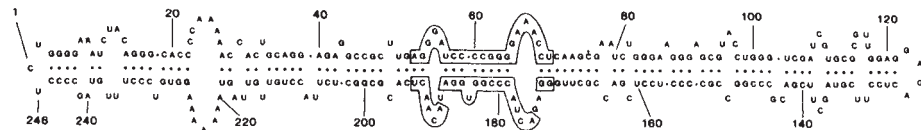
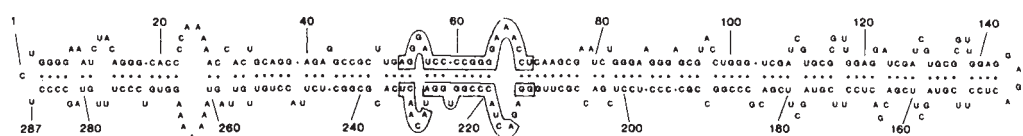


Fig. 1 Incidence of cadang-cadang disease of coconuts. Regions of the Philippine islands where the disease is widespread are shown cross-hatched, while isolated incidences of the disease occur in surrounding areas³². The different ccRNA isolates used in this work were each obtained from separate diseased coconut palms in one of the following locations: 1, San Miguel Island; 2, Sorsogon; 3, Ligao; 4, Lake Baao; 5, Tinambac; 6, Guinayangan; and 7, San Nasciso.

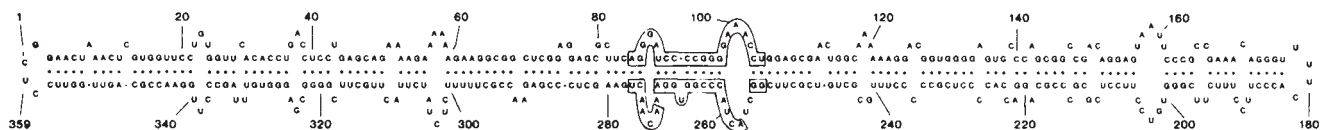
Cadang-cadang RNA 1 fast



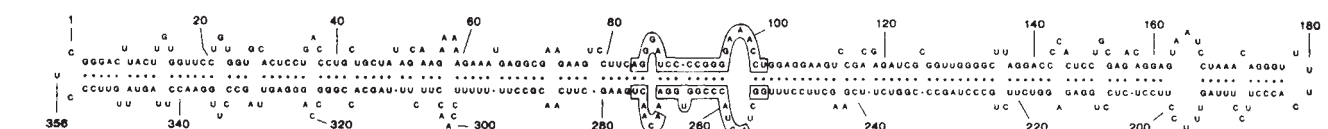
Cadang-cadang RNA 1 slow



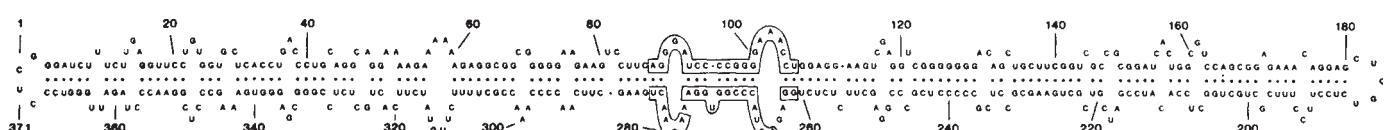
PSTV



CSV



CEV



ASBV

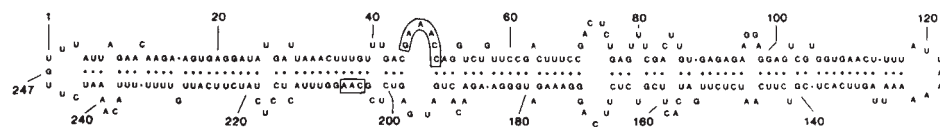


Fig. 2 Sequences and predicted secondary structures of the Baoo 54 isolate of ccRNA 1 fast and ccRNA 1 slow are shown with those of PSTV¹⁸, CSV¹⁰, CEV¹² and ASBV¹¹. The structures are aligned under the central conserved regions of these viroids (boxed). For sequence determination, nucleic acids were extracted from the fronds of a single infected coconut palm³³, and ccRNAs purified by preparative polyacrylamide gel electrophoresis⁴. Purified ccRNAs were sequenced essentially as described in ref. 10. Thus, 5 µg circular ccRNA in 20 µl 600 mM NaCl, 10 mM MgCl₂, at 0°C was digested with RNase T₁ (Sigma) at 2,000 units ml⁻¹ to produce specific full-length linear molecules, or with 3 units ml⁻¹ RNase U₂ (Sankyo) or 5 ng ml⁻¹ RNase A (Sigma) to produce smaller linear fragments. These fragments were 5'-radiolabelled using [γ-³²P]ATP and T4 polynucleotide kinase or, after treatment with calf intestinal phosphatase³⁴, 3'-radiolabelled using [3'-³²P]dCp and T4 RNA ligase¹⁴, and fractionated on a 80×20×0.05 cm 6% polyacrylamide gel containing 8 M urea and TBE buffer (90 mM Tris-borate pH 8.3, 2 mM EDTA). Bands located by autoradiography were excised and eluted¹⁰, and sequenced using partial enzymatic cleavage¹⁰. The sequences of numerous overlapping fragments were assembled to give the complete primary structure of each circular molecule. For S₁ nuclease mapping of the secondary structures⁷, 5'- or 3'-radiolabelled full-length linear fragments, obtained as described above by RNase T₁ digestion, were suspended in 20 µl 200 mM NaCl, 0.5 mM ZnSO₄, 50 mM sodium acetate pH 4.6, containing 5 µg *Escherichia coli* tRNA carrier, and incubated at 37°C for 10 min with 0.1, 1 or 10 units of S₁ nuclease (Boehringer). The reaction mixture was extracted with phenol, the RNA precipitated with ethanol and fractionated by polyacrylamide gel electrophoresis in TBE, 8 M urea. Products of partial enzymatic sequencing reactions of the same ccRNA were run as markers, thus allowing sites of S₁ nuclease sensitivity to be located. Data so derived were used to construct the secondary structure models of ccRNA 1 fast and ccRNA 1 slow.

of each ccRNA 2 species produced specific linear RNA fragments corresponding to both the respective full-length ccRNA 2 and ccRNA 1 molecules. Sequence determination of these fragments showed that cleavage of the ccRNA 2 molecules occurred at two sites located at the same sequences as for the two ccRNA 1 molecules. This suggests that either the predicted conformer A of the two ccRNA 2 molecules or possible cruciform intermediates exist in solution whereby the terminal hairpin loops are exposed. However, the existence of type B conformers cannot be precluded.

Variation in sequence between different ccRNA isolates

Different isolates of cadang-cadang RNAs were each obtained from single infected coconut palms from different localities in the Philippines (Fig. 1). Sequence differences between the different isolates consist of two types. First, the sequences of

the ccRNA 1 slow forms can differ. While all ccRNA 1 fast forms are essentially identical (see below), ccRNA 1 slow forms can differ in the length of the repeated sequence inserted between residues 123 and 124 of ccRNA 1 fast. Three different repeated sequences found in nine sequenced isolates of ccRNA 1 slow are given in Fig. 4; these vary in length from 41 to 55 residues but they are internally base-paired to produce duplicated structures as well as sequences at the right-hand ends of the molecules. Interestingly, the right-hand ends of the native molecules of PSTV, CSV, CEV and ASBV (Fig. 2) are at a similar distance from the central conserved regions of these molecules. In contrast, the right-hand side of the ccRNA 1 fast molecule is shorter while those of the elongated ccRNA 1 slow molecules are closer in size to those of PSTV, CEV and ASBV.

Second, four of the six isolates of ccRNA 1 fast sequenced each consist of two populations of molecules, one of 246 residues and the other of 247 residues, which differ in the

presence or absence of a C at residue 198 (Fig. 5, Table 1). Similar sequence heterogeneities have also been reported for CEV¹⁹ and the viroid-like RNA of VTMOV¹³. The relative proportions of the two ccRNA 1 fast subspecies vary between different isolates as listed in Table 1. For the two ccRNA 2 fast isolates sequenced, the relative proportions of the two forms are the same as those of the corresponding ccRNA 1 fast. In contrast to the ccRNA 1 and 2 fast species, no similar sequence heterogeneity has been observed in nine isolates of ccRNA 1 slow and the one sequenced isolate of ccRNA 2 slow (Table 1). Each isolate of the ccRNA slow species thus consists entirely of either one subspecies or the other; in all except one case, the C at the position corresponding to ccRNA 1 fast residue 198 was absent. The various sequence differences between the ccRNA isolates do not seem to correlate with differences in geographical location.

Structural similarities between ccRNAs, viroids and virusoids

The ccRNAs share two regions of sequence homology, each of about 20 nucleotides, with the viroids PSTV, CSV and CEV

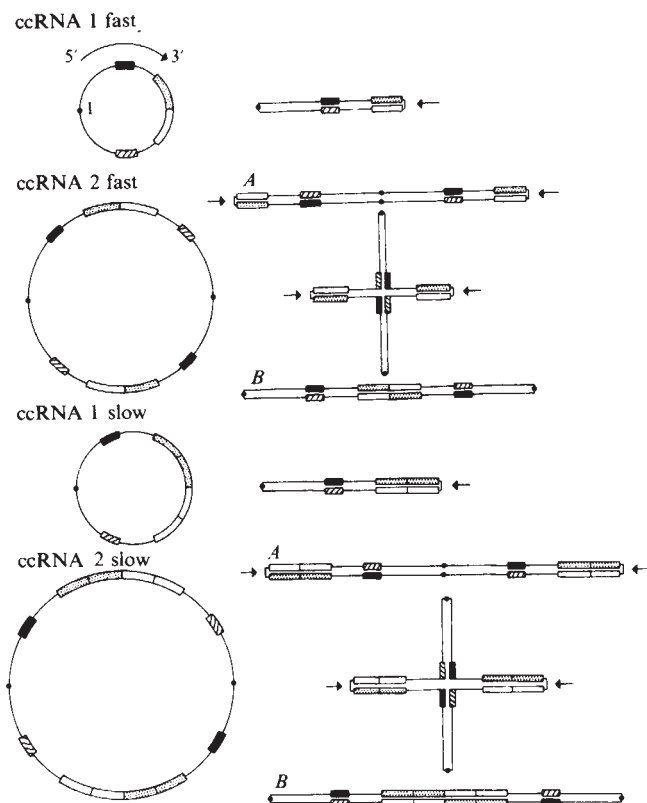


Fig. 3 Schematic representation of the sequences and predicted structure relationships between the ccRNAs. The circular sequences of the four ccRNAs are shown with black boxed and cross-hatched boxed regions representing the sequences highly conserved between the ccRNAs, PSTV, CSV and CEV (Baao 54 ccRNA 1 fast residues 52–71 and 171–194, respectively). The white boxed and stippled boxed regions represented those sequences duplicated within the ccRNA 1 slow species (Baao 54 ccRNA 1 fast residues 103–123 and 124–143, respectively). Positions corresponding to residue 1 of ccRNA 1 fast or ccRNA 1 slow are indicated by black dots. Both ccRNA 2 fast and ccRNA 2 slow molecules are dimers of their respective ccRNA 1 forms. Each ccRNA 2 can potentially form either of two rod-like conformers, A or B, as well as a large number of cruciform-shaped intermediates of which one is shown for ccRNA 2 fast and ccRNA 2 slow. Each ccRNA 1 species possesses a single, highly accessible site for RNase T₁ cleavage located on a hairpin loop (between Baao 54 ccRNA 1 fast residues 124 and 125 and between ccRNA 1 slow residues 145 and 146); these sites are indicated by arrows. Each ccRNA 2 species possesses two such accessible sites for RNase T₁ cleavage and arrows indicate where these sites also occur on hairpin loops in the different ccRNA 2 conformers.

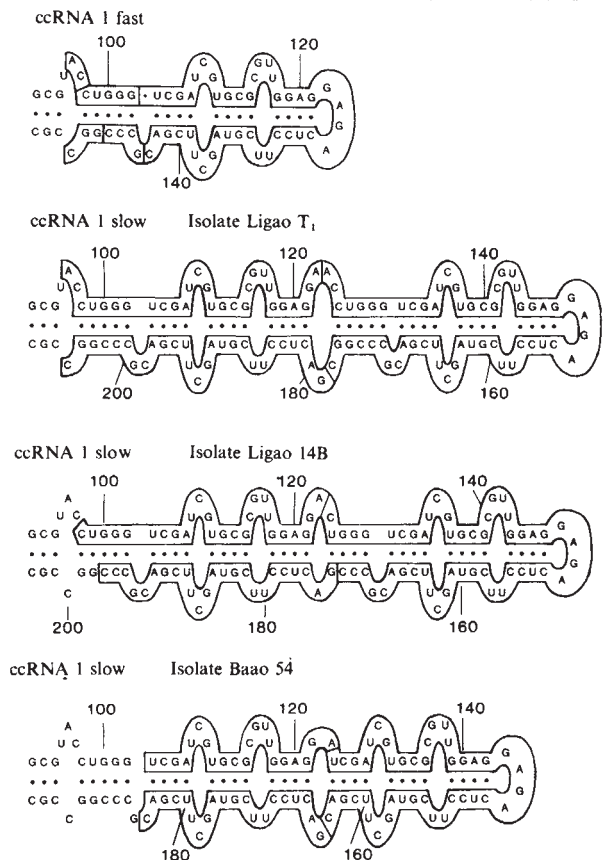


Fig. 4 Sequence variation between ccRNA 1 slow of three ccRNA isolates. The sequences and structures of various ccRNA 1 fast and ccRNA 1 slow isolates were determined as described in Fig. 2 legend. As essentially all sequence variation occurred at the right-hand end of the ccRNA 1 slow molecules, only this region is shown. Boxed regions represent those sequences which are duplicated in the ccRNA 1 slow molecules and which are 41 (isolate Baao 54), 50 (isolate Ligao 14B) or 55 residues (isolate Ligao T₁) long. All sequenced ccRNA 1 slow isolates correspond to one of these forms (Table 1).

(Fig. 2). The latter three viroids are closely related, sharing about 50% sequence homology. These two conserved regions are base-paired in the predicted structures of the native molecules to form highly conserved secondary structures.

The conserved regions shared by the ccRNAs correspond to regions of PSTV, CSV and CEV postulated to be involved in base pairing of viroid complementary RNAs with a plant small nuclear RNA (snRNA) in a manner analogous to that proposed for the interaction of mRNA intron-exon splice junctions with mammalian Ula snRNA^{19,20}, and in the formation of a stabilizing stem-loop structure in the viroid complements¹⁹. The proposed interaction between viroid complements and snRNA is postulated to reflect the origin of viroids from an intron ancestor²⁰ or as a basis for pathogenesis^{19,20} but not to be directly involved in viroid replication. However, the complementary RNA of at least one viroid, ASBV, is incapable of base pairing with a Ula-like snRNA or of formation of the conserved stem-loop structure, despite up to 18% sequence homology between ASBV and other viroids¹¹. It is possible that the central conserved regions of native viroids reflect functional similarities related to viroid replication rather than to the postulated snRNA binding.

Interestingly, the virusoids (encapsidated viroid-like RNAs) of VTMOV¹³, SNMV¹³, and subterranean clover mottle virus (J.H., C. Davies and R.H.S., unpublished data) also contain highly conserved sequences which are central to their rod-like native structures and which share only the pentanucleotide sequence, GAAAC, with all sequenced viroids, including ASBV.

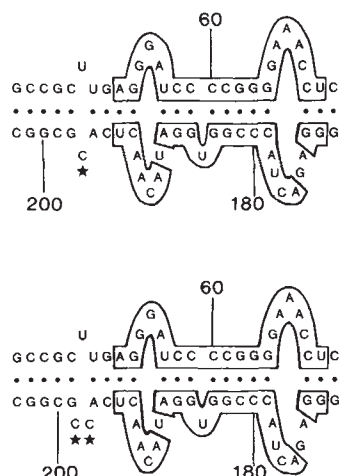


Fig. 5 Portions of the predicted native structures for the two forms of ccRNA 1 fast molecules. Purified ccRNA isolates were found to consist of either one or a mixture of two RNA species which differed in the presence or absence of an extra C residue at the position corresponding to ccRNA 1 fast residue 197 or 198. Stars indicate the sequence differences. Boxed regions indicate those sequences common to PSTV, CSV and CEV.

Replication of ccRNAs

As PSTV, CSV, CEV and the ccRNAs are capable of autonomous replication, the enzymes involved are of considerable interest. However, no viroid-encoded translation products have been found *in vitro*^{21,22} or *in vivo*²³. Although PSTV, CSV and CEV share ~50% sequence homology, none of these viroids nor their putative complementary RNAs can theoretically encode similar translation products^{10,12}, even assuming the existence of translatable linear viroid RNAs *in vivo*^{24,25}. Possible protein-coding regions similar to those of other viroids are not found in the ccRNAs or their complements nor are there any AUG initiation codons present. As a 5'-proximal AUG initiation codon seems essential for initiation of translation of eukaryotic RNA^{26,27}, it seems highly unlikely that the ccRNAs can code for any functional polypeptide product. All evidence, therefore, indicates that ccRNAs and other viroids must rely entirely on host components for their replication.

Larger than unit-length complementary (-) RNA intermediates have been detected in PSTV- and CEV-infected tissues²⁸⁻³⁰, and recently an oligomeric series of RNAs of ASBV (+) have been detected in infected avocado tissue (G. Bruening, A. R. Gould, P. J. Murphy and R.H.S., unpublished results). Rolling circle mechanisms have been postulated for the synthesis of oligomeric complementary RNAs from circular viroid templates^{28,30}, and oligomeric viroid (+) sequences could be simply generated by transcription of multimeric (-) strand templates. Unit-length linear viroid produced by either specific transcription or cleavage of oligomeric viroid RNAs must be ligated to produce the final circular product. Such a model for viroid replication, involving oligomeric RNA intermediates, could readily account for the formation of the dimeric forms of both ccRNA 1 fast and ccRNA 1 slow; that is, ccRNA 2 fast and ccRNA 2 slow, respectively. Rate-limiting steps during the transcription or the possible processing of viroid transcripts would allow the dimeric ccRNAs 2 to accumulate over the

monomeric ccRNA 1 species. The presence or absence of higher multimers of (+) ccRNAs or the complementary (-) ccRNAs in infected tissue has yet to be determined.

ccRNA slow variants and the time course of infection

In the initial stages of cadang-cadang disease, only the fast forms of ccRNA 1 and ccRNA 2 are present in infected palms and it is only after a further 24-30 months that the slow variants of ccRNA 1 and ccRNA 2 first appear and in the following years predominate⁵. These data, plus preliminary evidence that the ccRNA fast species are more infectious than the ccRNA slow species⁴, are consistent with the *de novo* generation of the ccRNA slow variants during each cadang-cadang disease infection. This proposition is supported by the following sequence data.

- (1) The ccRNA 1 slow forms differ from ccRNA 1 fast by the insertion of a single repeated sequence (Fig. 2) and could be simply generated from the ccRNA 1 fast by processing and/or transcription mechanisms.
- (2) The ccRNA 1 slow isolates can differ in the size of their inserted sequence repeats (Fig. 4, Table 1) suggesting separate origins for these ccRNA slow variants.
- (3) While most ccRNA fast isolates contain a sequence heterogeneity at residue 198, and consist of varying ratios of the 246- and 247-residue species, each of the nine sequenced ccRNA slow isolates consists of a single homogeneous population, either with or without a C residue at the position homologous to ccRNA 1 fast residue 198, and with only one size of repeated sequence (Table 1).

These data are consistent with the generation of ccRNA slow forms from ccRNA fast by single, rare sequence duplication events occurring separately in each cadang-cadang-infected palm. All ccRNA slow molecules would, therefore, originate from single parent molecules and may accumulate in preference to ccRNA fast species due to a competitive advantage in replication.

Origin of cadang-cadang disease

The ccRNAs share biological properties and sequence and structural homology with viroids so that application of the term coconut cadang-cadang viroid is fully justified. However, whereas other viroids consist of a single predominant infectious RNA species, CCCV consists of several variant RNA species. It is feasible that CCCV may have arisen from a pre-existing viroid and that mutation or infection of new hosts, such as the coconut palm and related species³¹, resulted in production of the variant ccRNAs by aberrant transcription and/or processing mechanisms which normally occur faithfully in the replication of other viroids. The outbreak and subsequent apparent rapid spread of cadang-cadang disease in the Philippines this century³² is consistent with such an origin of the ccRNAs. Future work with coconut palms inoculated with purified ccRNA species should allow a study of the generation of the ccRNA variants and may help elucidate the nature of viroid replication.

This work was supported by the Australian Research Grants Scheme and the UNDP/FAO Coconut Research Project. We thank Dr B. Zelazny for advice, Dr D. Riesner for unpublished data, Dr A. J. Gibbs for suggesting the term virusoids for encapsidated viroid-like RNAs, and Jenny Rosey, Lisa Waters and Julita Imperial for assistance.

Received 28 June; accepted 16 August 1982.

1. Ocfemia, G. P. *Philipp. Agric.* **26**, 338-340 (1937).
2. Bigornia, A. E. & Philipp, J. *Coconut Stud.* **2**, 5-33 (1977).
3. Boccardo, G., Beaver, R. G., Randles, J. W. & Imperial, J. S. *Phytopathology* **71**, 1104-1107 (1981).
4. Imperial, J. S., Rodriguez, J. B. & Randles, J. W. *J. gen. Virol.* **56**, 77-85 (1981).
5. Mohamed, N. A., Haseloff, J., Imperial, J. S. & Symons, R. H. *J. gen. Virol.* (in the press).
6. Mohamed, N. A. & Imperial, J. S. (in preparation).
7. Randles, J. W., Rillo, E. P. & Diener, T. O. *Virology* **74**, 128-139 (1976).
8. Randles, J. W. & Hatt, T. *Virology* **96**, 47-53 (1979).
9. Gross, H. J. & Riesner, D. *Angew. Chem. int. Edn.* **19**, 231-243 (1980).

10. Haseloff, J. & Symons, R. H. *Nucleic Acids Res.* **9**, 2741-2752 (1981).
11. Symons, R. H. *Nucleic Acids Res.* **9**, 6527-6537 (1981).
12. Visvader, J. E., Gould, A. R., Bruening, G. E. & Symons, R. H. *FEBS Lett.* **137**, 288-292 (1982).
13. Haseloff, J. & Symons, R. H. *Nucleic Acids Res.* **12**, 3681-3691 (1982).
14. England, T. E. & Uhlenbeck, O. C. *Nature* **275**, 560-561 (1978).
15. Kramer, F. R. & Mills, D. R. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5334-5338 (1978).
16. Tinoco, I., Uhlenbeck, O. C. & Levine, M. D. *Nature* **230**, 362-367 (1971).
17. Wurst, R. M., Vournakis, J. N. & Maxam, A. M. *Biochemistry* **17**, 4493-4499 (1978).
18. Gross, H. J. *et al. Nature* **273**, 203-208 (1978).
19. Gross, H. J. *et al. Eur. J. Biochem.* **121**, 249-257 (1982).
20. Diener, T. O. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5014-5015 (1981).

21. Davies, J. W., Kaesberg, P. & Diener, T. O. *Virology* **61**, 281–286 (1974).
22. Semancik, J. S., Conjero, V. & Gerhart, J. *Virology* **80**, 218–221 (1977).
23. Conjero, V. & Semancik, J. S. *Virology* **77**, 221–232 (1977).
24. Kozak, M. *Nature* **280**, 82–85 (1979).
25. Konarska, M., Filipowicz, W., Domdey, H. & Gross, H. J. *Eur. J. Biochem.* **114**, 221–227 (1981).
26. Baralle, F. E. & Brownlee, G. G. *Nature* **274**, 84–87 (1978).
27. Kozak, M. *Cell* **15**, 1109–1123 (1978).

28. Branch, A. D., Robertson, H. D. & Dickson, E. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6381–6385 (1981).
29. Rohde, W. & Sanger, H. L. *Biosci. Rep.* **1**, 327–336 (1981).
30. Owens, R. A. & Diener, T. O. *Proc. natn. Acad. Sci. U.S.A.* **79**, 113–117 (1982).
31. Randles, J. W., Boccardo, G. & Imperial, J. S. *Phytopathology* **70**, 185–189 (1980).
32. Zelazny, B. *Acta phytopathol. Acad. Sci. Hung.* **14**, 115–126 (1979).
33. Randles, J. W. *Phytopathology* **65**, 163–167 (1975).
34. Efstratiadis, A. *et al. Nucleic Acids Res.* **4**, 4165–4174 (1977).

LETTERS TO NATURE

Emission mechanism and source distances of γ -ray bursts

Edison P. Liang

Lawrence Livermore National Laboratory, University of California, Livermore, California 94550, USA and *Institute for Plasma Research, Stanford University, Stanford, California 94305, USA

The reported continuum spectra (~ 20 keV–2 MeV) of most γ -ray bursts resemble a simple exponential^{1–4}. Hence, the conventional view is that they are the optically thin free-free (bremsstrahlung) emission of a hot thermal plasma ($kT \geq 100$ keV with the exception of a few soft events, notably the 5 March 1979 event). I show here that: (1) independent of the source distance, there is an inherent difficulty with the free-free interpretation if the sources are indeed neutron stars with magnetic field $B \geq 10^{12}$ G, because the synchrotron emissivity of these hot thermal electrons would greatly exceed their free-free emissivity for any reasonable electron density consistent with $\tau_{\text{es}} \ll 1$ required by observations; (2) the spectral data can be fitted equally well, if not better, by optically thin thermal synchrotron spectra of mildly relativistic ($kT \sim mc^2$) electrons; (3) the absence of observable low-energy turnover due to synchrotron self-absorption (or high energy Compton distortion) puts interesting upper limits to the source of luminosity and distance. Most sources are noncosmological (< 10 kpc–1 Mpc) but not necessarily as nearby as some authors have suggested using the free-free spectral interpretation (see, for example, ref. 2).

The essential arguments only are outlined here and illustrated with a few examples. Systematic analyses of a large collection of reported γ bursts have been performed and will be reported elsewhere⁵. Throughout this letter $T \equiv kT_e/mc^2$ (T_e = electron temperature; m = electron mass); $\nu_L \equiv eB/2\pi mc = 11.6$ keV ($B/10^{12}$ G) and $\nu_c \equiv \nu_L T^2$.

In the optically thin limit, the total free-free luminosity is given by⁶

$$L_{\text{ff}} \approx 1.2 \times 10^{-22} n_e^2 V Z T^{1/2} \text{ erg s}^{-1} \quad (1)$$

where n_e is electron density, Z is atomic number and V is emission volume, while the synchrotron luminosity from a thermal plasma is⁷ (compare with refs 8, 9)

$$L_{\text{syn}} \approx 2.7 \times 10^9 n_e V T^2 \mathcal{F}(T) B_{12}^2 \text{ erg s}^{-1} \quad (2)$$

where $\mathcal{F} = T^{-1}$ for $T \ll 1$; $\mathcal{F} = T^2/K_2(1/T)$ for $T \geq 1$, (K_2 = Bessel function); $B_{12} = B/10^{12}$ G. Hence, $L_{\text{ff}} \geq L_{\text{syn}}$ only if

$$n_e \geq 2.3 \times 10^{31} T^{3/2} \mathcal{F}(T) B_{12}^2 Z^{-1} \text{ cm}^{-3} \quad (3)$$

independent of source distance. The absence of any Compton tail or 'Wien hump' implies that $\tau_{\text{es}} = \sigma_{\text{es}} n_e h \ll 1$ where σ_{es} is the (Klein–Nishina) Compton cross-section and h is the thickness of the emission region. Hence, equation (3) requires

$$h \ll 2 \times 10^{-7} T^{-3/2} \mathcal{F}^{-1} B_{12}^{-2} Z \text{ cm} \quad (4)$$

Since $T \sim 1$ for most bursts¹, h is inconceivably small even for fields much below 10^{12} G, keeping in mind that typical bursts

last longer than seconds. (The gravitational scale height is, of course, huge: $h_G \sim 3.6 \times 10^3 T$ cm, but even magnetic confinement has difficulty because equation (3) says that gas pressure often exceeds magnetic stress: $P_{\text{gas}}/P_{\text{mag}} \geq 10^3 T^{5/2} Z^{-1}$. The above difficulty is actually much more acute if we concentrate on the specific emissivities of the observed frequency range (20 keV–2 MeV) instead of the total emitted power. This is because much of the free-free emission contained in equation (1) comes out at X-ray energies, whereas the synchrotron spectrum has more power in γ -ray energies (compare with equation (5)).

There is now widespread belief that most γ -ray bursts originate from neutron stars with strong magnetic fields (see ref. 10 for review). The discovery of cyclotron absorption lines by Mazets *et al.*¹, if they can be confirmed, would also give support to the presence of strong fields in the sources. The above arguments thus force us seriously to consider thermal synchrotron as the universal emission mechanism of γ -ray bursts. (See refs 11, 12 on the 5 March event. This has also been emphasized by Lamb¹⁰ and Katz¹³.)

The specific synchrotron emissivity of mildly relativistic electrons ($T \sim 1$) assume a simple analytical form in the regime $\nu \gg \nu_{L/T}$ (refs 7, 8)

$$j_\nu(\theta) = \left(\frac{2\pi e^2}{3c} \right) \nu n_e \cdot K_2^{-1}(1/T) \cdot \exp(-(4.5\nu/\nu_c \sin \theta)^{1/3}) \quad (5)$$

Hence, the photon number spectrum has the form

$$n_\nu \text{ (photons cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1} \text{ keV}^{-1}) \propto \exp(-(4.5\nu/\nu_c \sin \theta)^{1/3}) \quad (6)$$

where θ is the angle between the observer direction and the (average) field direction. Figure 1 shows the spectral fits with equation (6) to several representative γ -ray spectra reported by Mazets *et al.*¹ and Gilman *et al.*². It is clear that the synchrotron fits are at least as good as the free-free exponential^{1,2}. In each case the fit gives the value of $\nu_c \propto B T^2$ (assuming that the flux is dominated by emissions near $\theta = \pi/2$) but no direct information on B or T separately. For most of the spectra we have studied, ν_c lies in the range 1–20 keV, with a clustering around 5–10 keV (see Table 1, also ref. 5). Because in almost all cases the rise of the continuum continues down to the detector limit, say 20–30 keV (ref. 1), we can only suppose that the cyclotron first harmonic lies below the detector limit, so that $B \leq 2\text{--}3 \times 10^{12}$ G, and $T \geq 0.2\text{--}1$ for the hot emission region. On the other hand, if the narrow absorption dips reported by Mazets *et al.*¹ are real and due to cyclotron absorption, there must exist colder ($kT \ll 100$ keV) regions with stronger fields ($B \sim 4\text{--}6 \times 10^{12}$ G) located between the continuum emission region and us. The asymmetry in many of Mazets' line profiles¹ with the absorption spreading towards the red (weaker fields in the cyclotron interpretation) may indeed suggest that the field is quite inhomogeneous, with the field strength increasing from the hot emission region to the (initially) cold absorption region. One may therefore visualize the 'expulsion' of the field by the hot emitting plasma (see ref. 14 for this scenario arising in the thermonuclear model). But how such a configuration can hold up for seconds is a significant problem.

It is reasonable to assume that the electrons can maintain a thermal (maxwellian) distribution in the first place? The synchrotron cooling time is extremely short:

$$t_{\text{syn cool}} \approx 4 \times 10^{-16} B_{12}^{-2} T^{-2} \mathcal{F} \text{ s}^{-1} \quad (7)$$

* Address for correspondence.

Table 1 Parameters corresponding to the spectra of Fig. 1

Spectra	ν_c (keV)	$\nu_{\min}$ (keV)	$T(\nu_L \leq \nu_{\min})$	τ_{es}^*	$L_{syn}^{*†}/A$	Flux [†] (observed)	d_{max}^* (kpc) $A^{1/2}$
a	10.1	20	≥ 0.71	8.6×10^{-4}	1.1×10^{41}	5.6×10^{-6}	12.6
b	4.6	30	≥ 0.39	4.3×10^{-4}	4.6×10^{40}	1.5×10^{-6}	15.9
c	0.9	20	≥ 0.21	5.8×10^{-5}	2.2×10^{39}	7.4×10^{-7}	5.0
d	5.8	50	≥ 0.34	5.3×10^{-4}	1.4×10^{41}	6.4×10^{-6}	13.5

Note that the small τ_{es} is consistent with the lack of a 'Wien hump'.

* Upper limits assuming $\nu_L \approx \nu_{\min}$, A is emission area in km^2 .

† Integrated over frequencies $\geq \nu_{\min}$, L_{syn} and flux are in erg s^{-1} and $\text{erg s}^{-1} \text{cm}^{-2}$, respectively.

and energy must be continuously supplied to the electrons on this time scale. Coulomb collisions are obviously too slow to thermalize the electrons:

$$t_{\text{coulb}} \sim 6 \times 10^{-12} T^{3/2} n_{24}^{-1} \text{ s} \quad n_{24} = n_e / 10^{24} \text{ cm}^{-3} \quad (8)$$

However, at least in laboratory situations it is widely believed that even collisionless plasmas can maintain near-maxwellian distributions via collective processes which operate on time scales of the order of plasma instability (for example, two-stream, beam-plasma, cyclotron resonance) growth times which are tied to the plasma or the cyclotron frequency¹⁵. Without delving into the controversy of whether such processes are indeed relevant, especially in strong fields, we only point out here that both the plasma oscillation time

$$\nu_p^{-1} \sim 10^{-16} n_{24}^{-1/2} \text{ s} \quad (9)$$

and electron gyration period

$$\nu_{\text{cyc}}^{-1} \sim 3 \times 10^{-19} B_{12}^{-1} \text{ s} \quad (10)$$

can be much shorter than $t_{\text{syn cool}}$. Thus a maxwellian electron distribution via collective processes is conceivable, at least for fields much below the quantum limit ($B_{\text{crit}} = 4.4 \times 10^{13} \text{ G}$).

The total synchrotron emissivity (ϵ) above a minimum frequency $\nu_{\min} > \nu_L/T$ can be obtained by integration of

equation (26) of ref. 7 (see also ref. 9) and is given approximately by the formula:

$$\epsilon = 2.7 \times 10^9 \text{ erg cm}^{-3} \text{ s}^{-1} n_e B_{12}^2 T^4 \cdot K_2^{-1}(1/T) e^{-x_m} (1 + x_m + \frac{x_m^2}{2!} + \frac{x_m^3}{3!} + \frac{x_m^4}{4!} + \frac{x_m^5}{5!}) \quad (11)$$

where $x_m \equiv (4.5 \nu_{\min}/\nu_c)^{1/3}$. On the other hand, the lack of synchrotron self-absorption turnover down to a frequency $\nu_{\min}$ means that (see ref. 9):

$$\frac{h\nu_{\min}}{m\nu_{\min}^2 T} \leq 1 \quad (12)$$

where we again assume $\langle \theta \rangle \approx \pi/2$ for the strongest constraint. Substituting the explicit form of j_ν (see equation (5)) we obtain

$$n_e h \leq 4.8 \times 10^{19} \nu_{\min} \text{ keV} T K_2(T^{-1}) e^{-x_m} \quad (13)$$

Since $L_{\text{syn}}(\nu \geq \nu_{\min}) = 10^{10} \cdot \epsilon h A$ where A is the emission surface area in units of km^2 , we obtain an upper bound to L_{syn} :

$$L_{\text{syn}} \leq 9.5 \times 10^{36} A T \nu_{\min} \text{ keV} \nu_c^2 \text{ keV} \left(1 + x_m + \frac{x_m^2}{2!} + \frac{x_m^3}{3!} + \frac{x_m^4}{4!} + \frac{x_m^5}{5!} \right) \text{ erg s}^{-1} \quad (14)$$

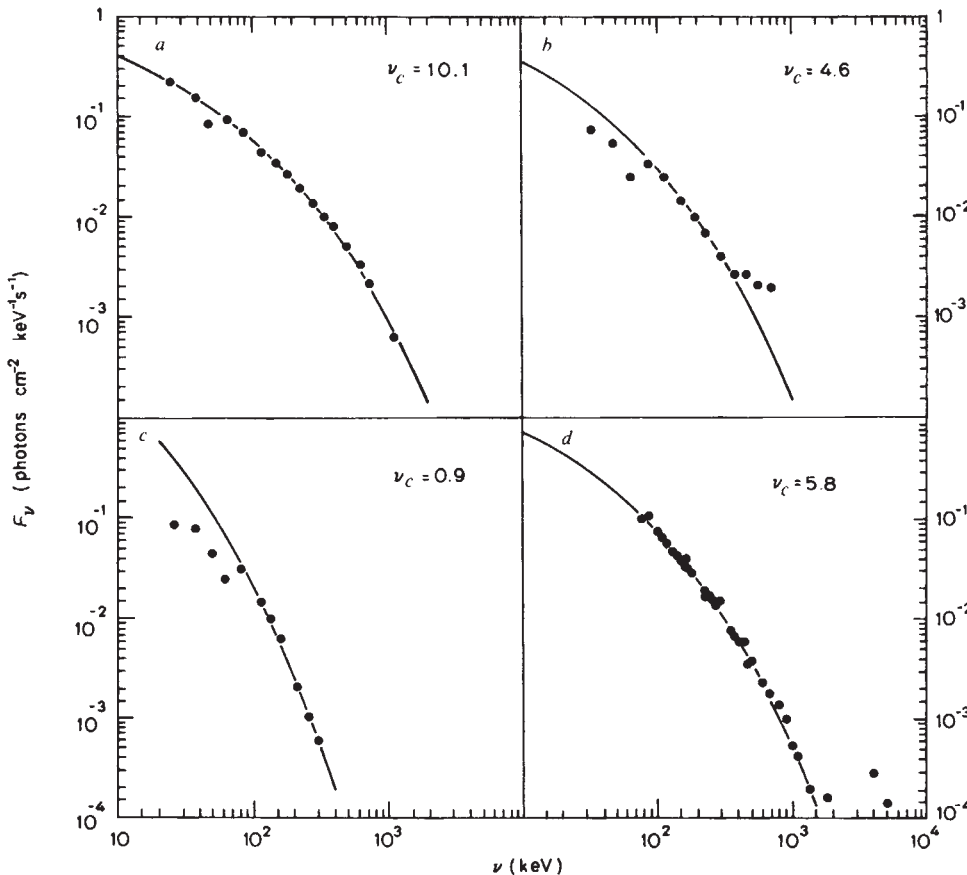


Fig. 1 Representative γ -ray burst spectra fitted with the thermal synchrotron formula (equation (6)). Spectra a, b correspond to Fig. 1a and 1b respectively of Mazets *et al.*¹. Spectrum c corresponds to Fig. 2d of Mazets *et al.*¹. Spectrum d corresponds to Fig. 3 of Gilman *et al.*². ν_c of the theoretical curves are in keV. The dips on the upper left of a, b, c were interpreted by Mazets *et al.*¹ as due to cyclotron absorption. The significance of the high energy excess is unclear because of huge error bars. For observational error bars and widths of energy bins readers should refer to the original articles^{1,2}.

For example, if we use the representative values of $\nu_{\min} = 20$ keV, $\nu_c \approx 10$ keV, and an observed flux of $\sim 6 \times 10^{-6}$ erg cm $^{-2}$ s $^{-1}$, then $x_m \approx 2$ and $L_{\text{syn}} \leq 1.5 \times 10^{41} T_A$ erg s $^{-1}$. If we further assume that $\nu_L \approx 20$ keV, then $T \approx 0.7$ and the upper bound to the intrinsic luminosity is $\sim 10^{41} A$ erg s $^{-1}$ and the source distance is $\leq 12 A^{1/2}$ kpc. Table 1 lists the upper bounds corresponding to the four examples of Fig. 1. Since L_{syn} decreases with $\nu_{\min}$ (see equation (14)), it is therefore important for future γ -burst observations to lower their detector energy limits to search for possible self-absorption turnover. We should also emphasize that equation (14) is almost definitely a lenient overestimate because in reality anisotropic pitch angle distributions will make the synchrotron luminosity less than that of an isotropic maxwellian plasma with the emission rate governed by pitch angle scattering rates.

We have discovered that optically thin thermal synchrotron from mildly relativistic electrons fits most observed γ burst spectra very well. The absence of self-absorption turnover down to 20–30 keV in most cases then suggests they lie within our local group of galaxies (≤ 1 Mpc). This, of course, does not contradict the suggestion from log N –log S analyses that most sources are likely galactic, since ours are upper limits which could be tightened by future observations at low energies. What is new is that these limits allow the possibility of some sources (for example, the 5 March event) being outside the galaxy. This

would not be possible using a free-free spectral interpretation (see refs 2, 16) since free-free luminosities are many orders of magnitude lower than synchrotron.

It has also been suggested recently that inverse Comptonization (of a soft photon source by hot electrons) may be the origin of the γ burst continuum¹⁷. While inverse Compton spectra may fit better than synchrotron or free-free in isolated cases, a preliminary survey of the Mazets catalogue shows that most of the spectra do not fit as well, since unsaturated inverse Compton spectra tend to have a more prominent 'knee' at $h\nu \sim 3kT_e$. More importantly, such fits would give lower electron temperatures ($T_e \leq 100$ keV) and a τ_{es} close to unity. For electrons in a 10^{12} G field, such large τ_{es} (compare with τ_{es} in Table 1) then leads to unacceptably large cyclotron luminosity. We would have to give up the strongly magnetized neutron star picture altogether.

We emphasize that no single spectral interpretation is binding in itself. But thermal synchrotron is physically more consistent with the conventional picture of strongly magnetized regions of neutron stars as the burst source. Confirmation of the synchrotron mechanism (for example, through the discovery of the first harmonics at low energies) will strongly affect future models of the bursts.

I thank Vahe Petrosian, Reuven Ramaty and George Chapline for useful discussions and Don Lamb for interesting comments.

Received 8 March; accepted 20 July 1982.

1. Mazets, E. P., Golenetskii, S. V., Aptekar', Guryan, Yu. A. & Ilinskii, V. N. *Nature* **290**, 378 (1981).
2. Gilman, D., Metzger, A. E., Parker, R. H., Evans, L. G. & Trombka, J. I. *Astrophys. J.* **236**, 951 (1980).
3. Cline, T. L. *Ann. N.Y. Acad. Sci.* **262**, 159 (1975).
4. Cline, T. L. & Desai, U. D. *Astrophys. J. Lett.* **196**, L43 (1975).
5. Liang, E. P. *Astrophys. J.* (submitted).
6. Tucker, W. *Radiation Processes in Astrophysics* (MIT Press, Cambridge, 1981).
7. Petrosian, V. *Astrophys. J.* (submitted).
8. Trubnikov, B. A. *Phys. Fluids* **4**, 195 (1961).

9. Bekefi, G. *Radiation Processes in Plasmas* (Wiley, New York, 1966).
10. Lamb, D. Q. *Proc. La Jolla Workshop on Gamma-Ray Transients and Related Astrophysical Phenomena* (ed.) Lingenfelter, R. (AIP, 1981).
11. Ramaty, R., Lingenfelter, R. E. & Bussard, R. W. *Astrophys. Space Sci.* **75**, 193 (1981).
12. Liang, E. P. T. *Nature* **292**, 319 (1981).
13. Katz, J. *Astrophys. J.* (submitted).
14. Woosley, S. E. & Wallace, R. K. *Astrophys. J.* (submitted).
15. Mikhailovskii, A. B. *Theory of Plasma Instabilities* (Consultants Bureau, New York, 1974).
16. Helfand, D. J. & Long, K. S. *Nature* **282**, 589 (1979).
17. Fenimore, E. E. et al. *Nature* **297**, 665 (1982).
18. Mazets, E. P. et al. *Astrophys. Space Sci.* **80**, 119 (1981).

Neutral hydrogen associated with the planetary nebula NGC6302

Luis F. Rodríguez

Instituto de Astronomía, UNAM, Apdo. Postal 70-264, 04510 Mexico DF, Mexico

James M. Moran

Harvard-Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, Massachusetts 02138, USA

Observations of the 21-cm line of neutral hydrogen in absorption made with the Very Large Array (VLA) towards the thermal radio source in the planetary nebula NGC6302 show two velocity components at 6 and -40 km s $^{-1}$ (radial velocity with respect to the local standard of rest). The 6 km s $^{-1}$ component is probably due to a line-of-sight cloud, while the -40 km s $^{-1}$ component is most likely associated with NGC6302. We interpret this latter absorption component as coming from the neutral, outer part of an expanding (~ 10 km s $^{-1}$) ring whose inner part is ionized and produces the absorbed thermal continuum radiation. The mass in atomic hydrogen of the outer (neutral) part of the ring is $\sim 0.06 M_{\odot}$. NGC6302 is probably in an evolutionary stage intermediate to those of protoplanetary nebulae such as GL2688 and evolved planetary nebula such as NGC7293. This is the first detection of H I associated with a planetary nebula.

Planetary nebula are believed to form from the expanding envelopes of red giants and supergiants¹. These stars have mass loss rates in the range 10^{-7} – $10^{-4} M_{\odot}$ yr $^{-1}$ with flow velocities of 10–20 km s $^{-1}$ (ref. 2). There is evidence that in many cases the outflow is not isotropic, but rather occurs preferentially in an equatorial plane³, presumably leading to a thick, expanding

disk around the star. It is not known how the final phases of mass loss occur. In a short time, which depends strongly on the stellar mass, the stellar radius decreases to only $\sim 0.02 R_{\odot}$ and the effective temperature of the star increases to $\sim 10^5$ K (ref. 4). The ionizing photons produced by the star will ionize the inner regions of the gaseous envelope. As the envelope expands and becomes less dense, it requires fewer UV photons to be ionized completely. Once complete ionization occurs, a classical planetary nebula, such as NGC6720 or NGC7293, formed by

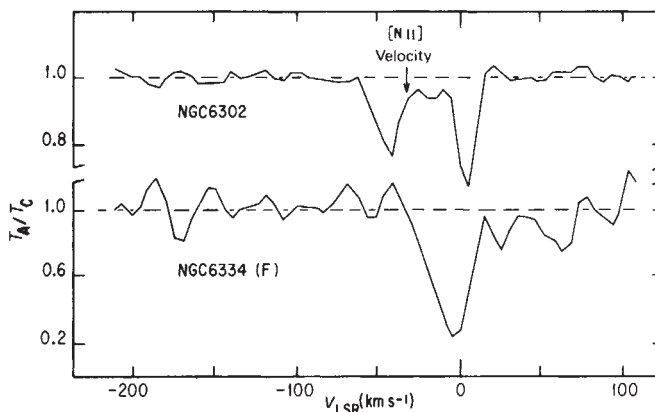


Fig. 1 H I absorption spectra towards the planetary nebula NGC6302 and the compact H II region NGC6334(F). The arrow marks the systemic velocity of NGC6302 from optical observations of [N II] (ref. 13). The absorption component at $V_{\text{LSR}} = -40$ km s $^{-1}$ is interpreted as coming from the outer region of an expanding ring around the planetary nebula. The horizontal axis is the radial velocity with respect to the local standard of rest and the vertical axis is the antenna temperature (T_A ; line plus continuum) over the continuum temperature T_c .

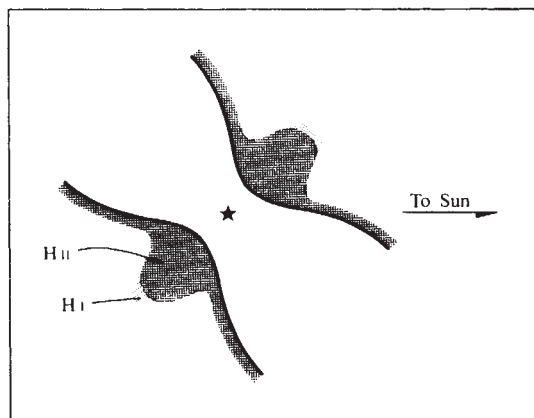


Fig. 2 A possible model for NGC6302. The central star is surrounded by an expanding ring of gas. The inner part of the ring is ionized and produces thermal continuum emission^{20,23}. The outer part of the ring is still neutral and causes the H I absorption that we observed.

an expanding ring of ionized gas, becomes visible. This last stage is the best studied and understood. Recently, considerable advances have been made in the understanding of the earlier evolutionary stages of forming planetary nebulae, mainly from radio and IR observations⁵. Particularly relevant to this understanding has been the identification of small nebulous objects as protoplanetary nebulae⁶⁻⁸, characterized by a predominance of molecular gas. When ionized gas is detected optically, it is found to originate from small regions with dimensions of $\sim 10^5$ cm or less, and consequently the radio flux is undetectable or weak^{9,10}. There should be objects between the protoplanetary nebula and the evolved planetary nebula stages which should have relatively strong thermal radio emission from the ionized gas, while retaining outer regions of neutral gas. A possible example is NGC7027, which is associated with a molecular cloud¹¹. However, the nature of this cloud is unclear as its mass exceeds $5 M_{\odot}$ (ref. 2), very large for a planetary nebula ($0.1-0.2 M_{\odot}$).

During 1982 February we made observations of the 21-cm line of H I in absorption against several compact radio sources in the molecular cloud complex NGC6334 (ref. 12) with the VLA of the National Radio Astronomy Observatory, located near Socorro, New Mexico. Details of the instrumental settings will be given elsewhere¹². We used 13 antennas of the VLA which was in the process of being moved from the C to A configuration. The angular resolution of the interferometer was ~ 10 arcs and the spectral window was ~ 330 km s⁻¹ with a resolution of 10.3 km s⁻¹. As a check source we observed NGC6302, which is only $\sim 2^{\circ}$ from NGC6334. NGC6302 has galactic coordinates of $l^{\text{II}} = 349^{\circ}$, $b^{\text{II}} = 1^{\circ}$ and we expected to detect absorption components only in the ± 10 km s⁻¹ range from local clouds. The observed spectrum is shown in Fig. 1. We detected an absorption component at $V_{\text{LSR}} = 6$ km s⁻¹, probably due to a line-of-sight cloud. A second component at -40 km s⁻¹ is also evident in the spectrum. This component is probably associated with NGC6302 for two reasons. First, the spectrum towards the nearby H II region NGC6334, also shown in Fig. 1, has no detectable absorption at -40 km s⁻¹. Absorption at such a velocity would be expected only from clouds farther than 5 kpc from the Sun. Second, the absorption feature is blueshifted by ~ 9 km s⁻¹ with respect to the radial velocity of the ionized gas in NGC6302, which is $V_{\text{LSR}} = -31 \pm 1$ km s⁻¹ (ref. 13). As the optical emission from the ionized gas originates from gas expanding symmetrically from the star, it can be considered to give a reliable estimate of the systemic velocity. Thus, the H I absorption appears to be due to gas that is expanding away from the nebula at 9 km s⁻¹. This expansion velocity is typical of planetary nebulae. In particular, the expansion velocity for NGC6302 derived from [O III] observa-

tions is 10 km s⁻¹ (ref. 14). However, the total dynamical picture is more complex. For example, higher velocity flows of ~ 100 km s⁻¹ have been seen in the ionized material¹⁵. To our knowledge this is the first detection of H I associated with a planetary nebula. Observing H I in absorption in planetary nebula is difficult as it requires the presence of a substantial amount of ionized gas to provide the background source and, at the same time, the presence of neutral gas. Probably, H I can be detected only during a transition phase, lasting $\sim 10^3$ yr, where the envelope changes from being neutral to being fully ionized. In 1970 Thompson and Colvin¹⁶ attempted to detect H I in six planetary nebulae, including NGC6302, but the sensitivity of their instrument was insufficient for this purpose. The presence of neutral gas in NGC6302 was already evident from the optical observations of Evans¹⁷, who detected strong [O I] line emission from it. Also, his short exposure photograph shows a dark lane separating the two lobes of NGC6302.

Our results can be combined with those of other authors^{15,17,18} to produce a model for NGC6302. Around the star there is a toroidal gaseous structure expanding at a velocity of ~ 10 km s⁻¹. The inner part of this toroid is ionized. Observations of the continuum of NGC6302 made with the VLA²⁰ show a ring-shaped structure with outer dimensions of $\sim 6 \times 9$ arcs. The outer part of the toroid is neutral (Evans' obscuring lane) and causes the H I absorption. The bipolar appearance of the nebula is due to the emission from ionized gas pushed out along the poles of the toroidal configuration. Figure 2 shows a schematic representation of this model, which is similar to that proposed by Canto *et al.*²¹ to explain bipolar outflows in pre-main-sequence objects.

Although the thermal source is ring-shaped with outer angular dimensions of 6×9 arcs, we will approximate it as a sphere of ~ 7 arcs diameter as this allows us to use the formulation of Milne and Aller²² to derive the distance of NGC6302. For a 6-cm flux of 2.8 Jy (refs. 20, 23), assuming the source is optically thin, a reasonable assumption based on the electron temperature, and an angular radius of ~ 3.5 arcs we obtain a distance of 2.4 kpc. Implicit in this derivation is the assumption that the mass in ionized hydrogen is $0.16 M_{\odot}$, the typical total mass of a planetary nebula. As we shall see, this assumption is reasonable since the mass in H I is smaller than the mass in H II. For an expansion velocity of 10 km s⁻¹ and a radius of $\sim 10^{17}$ cm (3.5 arcs at 2.4 kpc) we obtain a kinematic age of $\sim 4 \times 10^3$ yr for NGC6302.

Finally, we estimate the mass in H I. If we assume an excitation temperature of 100 K for the H I, then the column density is $N_{\text{H I}} \sim 8 \times 10^{20}$ cm⁻². If the H I covers the thermal continuum source, the mass of the foreground H I is $\sim 0.03 M_{\odot}$. Multiplying by 2 to account for the H I on the far side of the nebula gives $\sim 0.06 M_{\odot}$ for the total mass in H I. We ignore molecular hydrogen.

The several planetary nebulae which have characteristics similar to those of NGC6302 should be examined for H I. They probably are relatively young planetary nebulae with characteristics intermediate between those of protoplanetary nebulae and evolved planetary nebulae.

NRAO is operated by Associated Universities Inc., under contract with the NSF. We thank J. Meaburn and Y. Terzian for helpful comments.

Received 24 June; accepted 13 July 1982.

- Shklovsky, I. S. *Astr. Zh.*, **33**, 315-329 (1956).
- Knapp, G. R. *et al. Astrophys. J.* **252**, 616-634 (1982).
- Morris, M. *Astrophys. J.* **249**, 572-585 (1981).
- Paczynski, B. *IAU Symp.* **76**, 201-205 (1978).
- Zuckerman, B. A. *Rev. Astr. Astrophys.* **18**, 263-288 (1980).
- Zuckerman, B. *et al. Astrophys. J. Lett.* **205**, L15-19 (1980).
- Calvet, N. & Cohen, M. *Mon. Not. R. astr. Soc.* **182**, 687-704 (1978).
- Lo, K. Y. & Bechis, K. P. *Astrophys. J. Lett.* **205**, L21-25 (1976).
- Westbrook, W. E. *et al. Astrophys. J.* **202**, 407-414 (1975).
- Kwok, S. & Purton, C. R. *Astrophys. J.* **229**, 187-195 (1979).
- Mufson, S. L., Lyon, J. & Mariotti, P. A. *Astrophys. J. Lett.* **201**, L85-89 (1975).
- Moran, J. M., Rodríguez, L. F. & Backer, D. (in preparation).
- Minkowski, R. & Johnson, H. M. *Astrophys. J.* **148**, 659-662 (1967).
- Robinson, G. J., Reay, N. K. & Atherton, P. D. *Mon. Not. R. astr. Soc.* **199**, 649-657 (1982).

15. Meaburn, J. & Walsh, J. R. *Mon. Not. R. astr. Soc.* **193**, 631–640 (1980).
16. Thompson, A. R. & Colvin, R. S. *Astrophys. J.* **160**, 363–368 (1970).
17. Evans, D. S. *Mon. Not. R. astr. Soc.* **119**, 150–156 (1959).
18. Elliot, K. H. & Meaburn, J. *Mon. Not. R. astr. Soc.* **181**, 499–507 (1977).
19. Barral, J. F., Canto, J., Meaburn, J. & Walsh, J. R. *Mon. Not. R. astr. Soc.* **199**, 817–832 (1982).
20. Rodríguez, L. F. *et al.* (in preparation).
21. Canto, J., Rodríguez, L. F., Barral, J. F. & Carral, P. *Astrophys. J.* **244**, 102–114 (1981).
22. Milne, D. K. & Aller, L. H. *Astr. Astrophys.* **38**, 183–196 (1975).
23. Terzian, Y., Balick, B. & Bignell, C. *Astrophys. J.* **188**, 257–277 (1974).

Orientation of planetary O^+ fluxes and magnetic field lines in the Venus wake

H. Pérez-de-Tejada*, D. S. Intriligator†
& C. T. Russell‡

* Instituto de Geofísica, Universidad Nacional Autónoma de México, México 20 DF, México

† Carmel Research Center, PO Box 1732, Santa Monica, California 90406, USA

‡ Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024, USA

The presence of 'contaminant' heavy ions of planetary origin in the solar wind has long been the subject of intense theoretical and experimental research. Studies of their abundance, acceleration, and direction of motion are important because of their implications on the composition and dynamics of planetary and cometary plasma wakes. The plasma and magnetic field observations made with the Pioneer Venus Orbiter (PVO) at Venus have provided the opportunity to examine the conditions in which planetary ions are picked up by the solar wind. We show here that in the outer regions of the venusian far wake the displacement of planetary O^+ particles, characteristic of the Venus upper ionosphere, does not occur necessarily along the magnetic field lines but approximately in the direction of the shocked solar wind.

The mass loading of the solar wind plasma with atmospheric ions in non-magnetic planets was first examined in connection with the limited capacity of the flow to accommodate added mass¹. Contaminant material is expected to arise from photoionization processes and charge exchange collisions of the solar wind particles with exospheric atoms². In most studies^{3,4} the motion of the planetary ions is believed to result from electric $E \times B$ drifts superimposed on the Larmor gyration around the magnetic field lines. Wave-particle interactions associated with turbulent processes in the plasma^{5,6} may modify this motion, however, and produce a viscous-like acceleration of the planetary ions⁷.

The observation of heavy ions of planetary origin in the Venus wake was first reported⁸ from the low latitude orbital passes of the Venera 9 and 10 spacecraft. These ions were identified as a cool low-energy component directed into the umbra behind the terminator. The results of the spectral analyses of other Venera plasma data⁹ were also consistent with the presence of a component of planetary ions within the umbra.

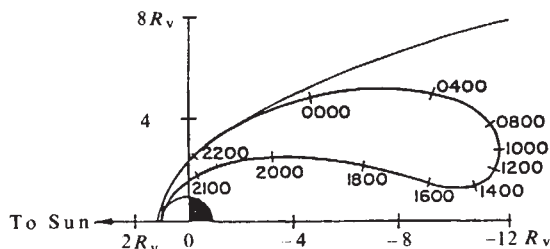


Fig. 1 Two-dimensional projection of the trajectory of the PVO in orbit 189, 11 June 1979.

The more recent measurements carried out with the PVO plasma probe^{10,11} have now revealed the detection of significant fluxes of O^+ ions of planetary origin in the Venus far wake. The energies of these ions are consistent with their having been accelerated to solar wind speeds, and their angular distributions imply that their direction of motion differs, at most, by only a few degrees ($\leq 5^\circ$) from that of the shocked solar wind. In addition, they seem to represent only a contaminant population because of the low ($\sim 1\%$) concentrations observed in the ambient shocked solar wind plasma. With such low densities their mass and momentum flux is significantly smaller than that of the shocked solar wind.

Simultaneous measurements of the magnetic field in the Venus wake carried out with the PVO magnetometer¹² have also indicated the existence of a peculiar magnetic geometry in that region. Most notable is a characteristic increase in the magnitude and fluctuation level of the magnetic field vector, which is predominately oriented along the Sun–Venus direction. Significant departures from this orientation may occur in the outer regions of the wake, however, and in particular at locations where the O^+ fluxes are detected.

From the comparison of the plasma and magnetic data in such regions, it is possible to examine the relative orientation of the magnetic field lines with respect to the direction of motion of the planetary heavy ions. Figure 1 shows the trajectory of the PVO during orbit 189 in a two-dimensional projection where the vertical coordinate is the distance perpendicular to the Sun–Venus axis. The spacecraft probed the magnetic wake between ~ 1230 and ~ 1730 UT and measured intense O^+ fluxes beginning at ~ 1210 UT. The aberrated azimuthal angle of these fluxes¹³ during the 1200–1400 UT time interval is illustrated in Fig. 2 together with the orientation of the local magnetic field (positive values result from directions with a component opposite to the orbital motion of the planet in a coordinate system where the x -axis is directed to the Sun).

The time interval shown is particularly useful because of the fairly persistent orientation of the magnetic field vector away from the solar direction between 1200 and 1300 UT. As shown in Fig. 2, this is predominately along directions which make appreciable negative angles with respect to that of the motion of the planetary ions. During this period the plasma probe measurements showed O^+ fluxes arriving from directions $\leq 5^\circ$ away from the x -axis, and thus with no apparent agreement with the orientation of the magnetic field lines. Simultaneous measurements of the direction of the shocked solar wind plasma showed that this differs at most by a few degrees from that of the O^+ fluxes (both traces would be, in fact, practically indistinguishable at the scale shown). A similar situation is observed between 1300 and 1400 UT despite the fact that during this time interval the magnetic field vector is significantly more variable. Thus, with the exception of the orientation seen near 1300 and 1350 UT, the magnetic field vector again makes large angles with respect to the direction of motion of the planetary ions. Further analysis of the data obtained later in this orbit also indicates significant differences between the direction of the plasma fluxes and the magnetic field orientation. Deeper within the wake, however, the observation of the O^+ particles is more sporadic and the magnetic field vector is seen to suffer frequent and sudden changes in magnitude and direction. In these conditions it is more difficult to establish a clear identification with respect to the direction of motion of the O^+ particles.

The overall tendency of the planetary ions to move in the direction of the shocked solar wind flow is indicative of the type of mechanism that should be responsible for their acceleration. Most notable is the fact that the direction of motion of the O^+ ions appears to be uncorrelated with changes in the direction of the magnetic field vector. Thus, the differences in the orientation of the particle and magnetic fluxes are important not only because they reveal that the displacement of the planetary particles is not necessarily along the field lines, but also because the observations show that the same drift velocity

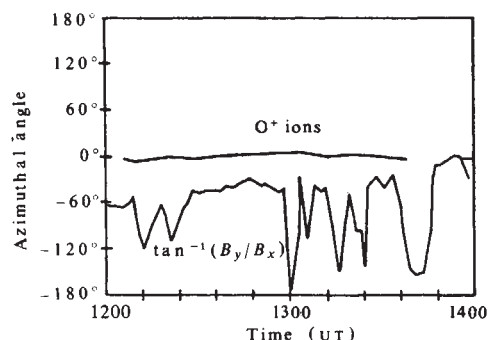


Fig. 2 Azimuthal angle of the magnetic field vector and of the direction of motion of the (peak intensity) O^+ fluxes in the Venus wake between 1200 and 1400 UT in orbit 189.

(nearly coincident with the shocked solar wind velocity) is measured consistently. This behaviour would not be expected if $E \times B$ drifts were solely responsible for the motion of the planetary O^+ ions. If this were the case, their direction of motion would reflect the changes of direction of the magnetic field encountered along the trajectory of the PVO. This circumstance seems to indicate that $E \times B$ pick up processes are not sufficient to account for the acceleration and direction of motion of the planetary ions.

As indicated above, the effects of wave-particle interactions associated with the turbulent character of the ionosheath flow, as revealed by the PVO plasma wave measurements¹⁴, should give place instead to stochastic motions, and enhance the fluid-like behaviour of the local plasma. In these conditions the acceleration process should result from collective interactions between the solar wind and the ionospheric particles, and produce an effective viscous drag between both particle populations⁷. This should set the O^+ particles into motion with speeds and directions nearly coincident with that of the streaming solar wind flow, and independent of the direction of the magnetic field. The PVO observation of heavy ion fluxes in this latter direction, and of the different orientation of the magnetic field vector, is consistent with that interpretation and indicates that wave-particle interactions should have an important role in the acceleration of the ionospheric particles.

As the interaction between the solar wind and the Venus environment is primarily with the atmosphere-ionosphere and not with a planetary magnetic field, the above results are also probably relevant to cometary plasma wakes. If this is the case, then downstream of a comet we should expect to observe accelerated cometary ions flowing in a direction similar to that of the solar wind and independent of the local orientation of the magnetic field.

Services and facilities provided by the Centro de Investigación y Educación Superior de Ensenada, Baja California, Mexico are gratefully appreciated. Partial support by NASA under research contracts: NAS2-10926 at the Carmel Research Center, NAS2-9491 at the University of California and A-80608B(DDA) for the PVO guest investigator programme is acknowledged.

Intergrown mica and montmorillonite in the Allende carbonaceous chondrite

Kazushige Tomeoka & Peter R. Buseck

Department of Geology and Chemistry, Arizona State University, Tempe, Arizona 85287, USA

Calcium-, aluminium-rich inclusions (CAIs) in carbonaceous chondrites such as Allende are of considerable interest because of their primitive character. There are, however, differences of opinions as to whether they formed directly by condensation from the early solar nebula¹ or by the evaporation of primitive dust². The mineralogy of the CAIs potentially bears on this question and therefore much effort has been devoted to detailed petrographic studies. We present here high-resolution transmission electron microscopy (HRTEM) observations of a complex mixture of mica and montmorillonite that occurs within a fine-grained CAI in the Allende C3(V) meteorite. In spite of previous intensive petrographic investigations of this and related meteorites, the existence of such CAIs bearing mica and montmorillonite has not been reported previously, and there have been no previous positive identifications of these minerals in carbonaceous chondrites.

We extracted part of an ordinary petrographic thin section that includes a CAI, and ion-thinned it for electron microscope observations. The inclusion, round and ~0.4 mm in largest dimension, consists of fine-grained fragments ranging in diameter from <1 to 50 μ m. Electron microprobe analyses show that the inclusion contains considerable Mg, Fe and Na in addition to major Si, Al and Ca, and that it exhibits extreme local compositional variations. The fine-grained fragments are mostly Fe-containing spinels that are characteristically enclosed by fine-grained rim material. These textural and chemical characteristics indicate that the inclusion corresponds to a fine-grained alkali-rich spinel aggregate in Wark's³ classification of CAIs and Type 3 in Kornacki's⁴ classification of fine-grained CAIs in Allende. The fine-grained CAIs are rich in alkali and volatile elements in addition to the refractory Ca-Al-rich minerals, which makes it difficult to explain their origin by simple condensation models⁵. Studies of petrographic characteristics of the fine-grained inclusions in the Allende meteorite are reported elsewhere^{3,6}.

In low magnification TEM images, we have observed phyllosilicate that shows a complex, fluffy texture. It occurs along two cavities that appear to be interstitial to spinel grains and are located close to the margin of the inclusion. Investigation by HRTEM has shown that much of the phyllosilicate has a (001) basal fringe spacing of ~10 Å (Fig. 1). This feature clearly distinguishes it from serpentine-type phyllosilicates that are now widely accepted as major constituents of the C2 matrix phases. Possible candidates for the phyllosilicates having such (001) spacings include mica, dehydrated montmorillonite-type clay minerals, and talc⁷.

The phyllosilicate grains display characteristic features in their morphologies and arrangements of lattice planes. Some crystals have a straight or sub-parallel arrangement of their lattice fringes. On the other hand, most crystals commonly show bending, terminating and overlapping fringes (Fig. 1). In some areas, fringes vary in spacing between 10 and 15 Å. These structural characteristics indicate a strong similarity to those observed in terrestrial mixed layered mica and montmorillonite⁸⁻¹⁰.

Chemical analyses were performed with an energy dispersive X-ray spectrometer (EDS) on our TEM and scanning electron microscope (SEM). EDS spectra from the phyllosilicate reveal large Si and Al peaks and less intense K, Mg, Ca and Fe peaks. Sodium was detected in significant but variable amounts, and minor Cl and Ti were also encountered. These results preclude the possibility of talc. Based on data from imaging and chemical analyses, we conclude that the phyllosilicate is a complex

Received 4 March; accepted 18 June 1982.

1. Michel, F. C. *Planet. Space Sci.* **15**, 19 (1971).
2. Wallis, M. K. & Ong, R. S. B. *Planet. Space Sci.* **23**, 713 (1975).
3. Cloutier, P. A., Daniell, R. E. Jr. & Butler, D. M. *Planet. Space Sci.* **22**, 967 (1974).
4. Cloutier, P. A. *Solar Wind Interaction with Planetary Ionospheres*, 111 (NASA SP-397, 1976).
5. Daniell, R. E. Jr. *J. geophys. Res.* **86**, 10094 (1981).
6. Curtis, S. A. *J. geophys. Res.* **86**, 4715 (1981).
7. Pérez-de-Tejada, H. *J. geophys. Res.* **85**, 7709 (1980).
8. Vaisberg, O. L. et al. *Physics of Solar Planetary Environments* vol. 2 (ed. Williams, D. J.) 904 (American Geophysical Union, Washington, DC, 1976).
9. Verigin, M. I. et al. *J. geophys. Res.* **83**, 3721 (1978).
10. Mihalov, J. D., Wolfe, J. H. & Intriligator, D. S. *J. geophys. Res.* **85**, 7613 (1980).
11. Intriligator, D. S. *Geophys. Res. Lett.* **9**, 727, 1982.
12. Russell, C. T., Luhmann, J. G., Elphic, R. C. & Scarf, F. L. *Geophys. Res. Lett.* **8**, 843 (1981).
13. Intriligator, D. S., Wolfe, J. H. & Mihalov, J. D. *IEEE Trans. Geosci. Remote Sensing* **GE-18**, 39 (1980).
14. Scarf, F. L., Taylor, W. L., Russell, C. T. & Elphic, R. C. *J. geophys. Res.* **85**, 7599 (1980).

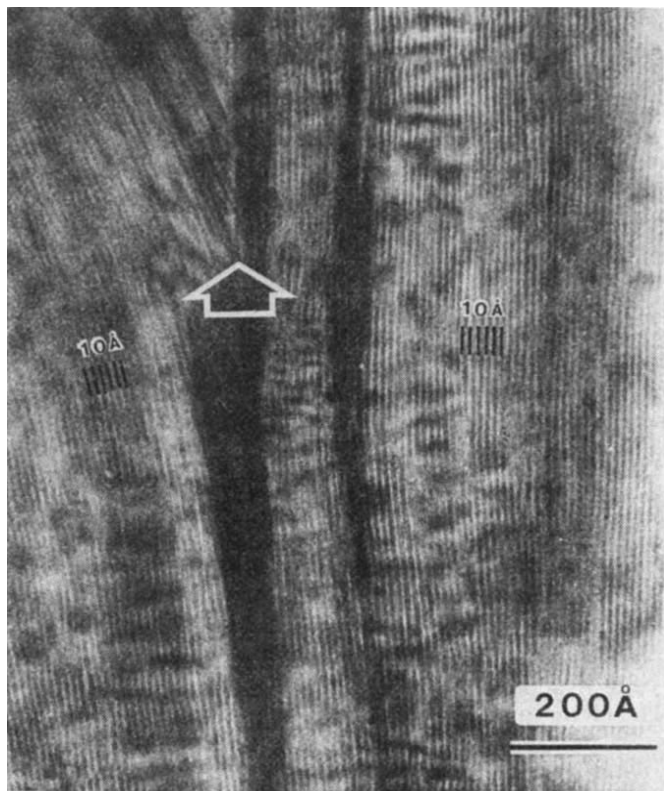


Fig. 1 An HRTEM image of an intergrown mica and montmorillonite crystal occurring along Ca-, Al-rich silicate grains within a fine-grained CAI of the Allende carbonaceous chondrite. The crystal exhibits bending, terminating and overlapping fringes (indicated by an arrow) and has a characteristic modulation in intensity along the length of the fringes.

mixture of mica and montmorillonite. Here the names are used as representing minerals having rather broad compositional ranges in these groups and not particular species.

The nature of the distribution of the Allende mica and montmorillonite along the walls of micro-cavities suggests that the phyllosilicate is genetically related to the surrounding alkali- and Ca-Al-rich silicate grains; owing to their fine-grained nature, the quality of chemical data from the surrounding crystals are not adequate to identify mineral species. They are probably feldspathic minerals, which constitute a rim material surrounding the spinel cores⁶. Recently, unusual unidentified phases have been reported to occur within rim material surrounding spinel grains in the fine-grained type of CAIs in Allende and Mokoia C3 meteorites^{11,12}; these phases have strikingly similar composition and texture to the present mica and montmorillonite assemblage. Aluminium-rich phyllosilicate was also reported from an inclusion within a C2-like clast in the Plainview (H5) meteorite¹³.

It is significant that the Allende mica and montmorillonite occur close to the boundary between the inclusion and matrix, and that the boundary is highly irregular in that region. Furthermore, a small spinel grain that is apparently related to the inclusion, occurs isolated within the matrix near the boundary. These observations suggest that reaction occurred between the inclusion and the matrix, and also that favourable conditions for producing the mica and montmorillonite existed near the boundary. McSween studied many C3(V) type meteorites and showed that there is apparent petrographic evidence of relatively mild metamorphism in several of those meteorites, and fine-grained olivine inclusions and CAIs in Allende may have undergone partial Fe/Mg exchange with matrix¹⁴. Our identification of mica and montmorillonite supports his view and suggests, as inferred from terrestrial analogues, that the metamorphism included aqueous alteration at a relatively low temperature.

Montmorillonite had been proposed as one of the possible phyllosilicate types that constitute a major portion of the matrix of C1 and C2 meteorites¹⁵⁻¹⁷. Recently, Barber reported fairly convincing evidence of the presence of montmorillonite in the matrix of the Nawapali C2(M) meteorite¹⁸. Bass suggested that montmorillonite replaces serpentine, providing for random mixed layering between them¹⁵. However, we did not encounter any evidence of serpentine intergrown with the mica and montmorillonite in Allende. On the contrary, the present results indicate that montmorillonite could be related to the Ca-Al-rich inclusions.

Armstrong *et al.*¹⁹ based on a detailed study of a hibonite-, calcite-bearing inclusion in the Murchison C2(M) meteorite, suggested that extensive alteration of CAIs may have occurred by aqueous alteration and thermal metamorphism before final meteorite formation. The scarcity of CAIs, especially the fine-grained type of inclusions in the C2 meteorites, could be explained by the susceptibility of CAIs to hydrothermal alteration. Oxygen isotopic data for fine-grained CAIs in the C2 meteorites would be useful for verifying whether or not such a hypothetical reaction had been widespread.

We thank John Larimer, Ian Mackinnon and David Veblen for helpful discussions and James Clark for assistance with electron microprobe analyses. The microprobe was obtained partly with funds provided by the Chemistry Instrumentation Program of the NSF. Electron microscopy was performed at the electron microscope facility in the Center for Solid State Science at Arizona State University. This work was supported by NASA grant NAGW-143.

Received 20 May; accepted 9 July 1982.

1. Grossman, L. *Geochim. cosmochim. Acta* **39**, 433-454 (1975).
2. Wood, J. A. *Earth planet. Sci. Lett.* **56**, 32-44 (1981).
3. Wark, D. A. *Astrophys. Space Sci.* **65**, 275-295 (1979).
4. Kornacki, A. S. *Lunar planet. Sci.* **XII**, 562-564 (1981).
5. Grossman, L. & Ganapathy, R. *Geochim. cosmochim. Acta* **40**, 967-977 (1976).
6. Kornacki, A. S. *Geochim. cosmochim. Acta* (submitted).
7. Veblen, D. R. & Buseck, P. R. *Science* **206**, 1398-1400 (1979).
8. McKee, T. R. & Buseck, P. R. *Electron Microscopy* **1**, 272-273 Intern. (1978).
9. Page, R. & Wenk, H. R. *Geology* **7**, 393-397 (1979).
10. *Kobutsugaku Zasshi Spec. Issue* **14**, 205-227 (Appendix) (1979).
11. Wark, D. A. & Lovering, J. F. *Proc. 8th Lunar Sci. Conf.* 95-112 (1977).
12. Cohen, R. E. *Meteoritics* **16**, 304 (1981).
13. Nozette, S. & Wilkening, L. L. *Geochim. cosmochim. Acta* **46**, 557-563 (1982).
14. McSween, H. Y. Jr *Geochim. cosmochim. Acta* **41**, 1777-1790 (1977).
15. Bass, M. N. *Geochim. cosmochim. Acta* **35**, 139-147 (1971).
16. Caillère, S. & Rautureau, M. C. *r. Acad. hebdom. Séanc. Sci., Paris* **D279**, 539-542 (1974).
17. Fanale, F. P. & Cannon, W. A. *Geochim. cosmochim. Acta* **38**, 453-470 (1974).
18. Barber, D. J. *Geochim. cosmochim. Acta* **45**, 945-970 (1981).
19. Armstrong, J. T., Meeker, G. P., Huneke, J. C. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **46**, 575-595 (1982).

An unusual layered mineral in chondrules and aggregates of the Allende carbonaceous chondrite

Kazushige Tomeoka & Peter R. Buseck

Departments of Geology and Chemistry, Arizona State University, Tempe, Arizona 85287, USA

Olivine-rich chondrules and aggregates in the Allende C3(V) carbonaceous chondrite are characterized by abundant opaque spherules of metal, sulphide and magnetite grains. This assemblage has been explained as resulting from the crystallization of immiscible metal-sulphide-oxide liquids, suspended as droplets within the chondrules¹. Petrographic evidence indicates a significant metasomatic effect in a highly oxidizing condition in a later cooled stage^{2,3}. We have found that an unusual Fe-, Ni- and O-rich layered mineral, related to serpentine, occurs within the opaque assemblage; it is formed by alteration of olivine. The textural and compositional features are clearly distinct from terrestrial phyllosilicates. This is the first report of a serpentine-related mineral in the Allende meteorite. Our observations suggest that aqueous conditions occurred in Allende before the final stage of meteorite formation.

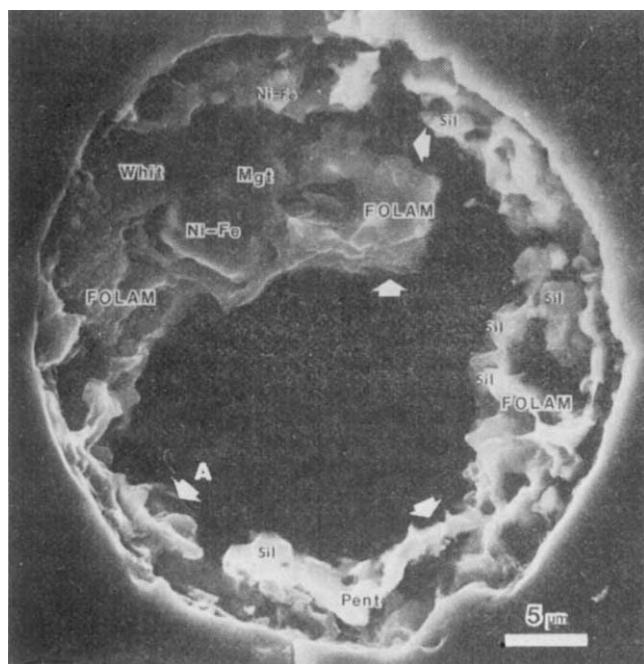


Fig. 1 SEM photograph of one of the ion-etched opaque spherules existing inside chondrules. The spherule contains Ni-Fe metal, magnetite (Mgt), pentlandite (Pent), whitlockite (Whit) and Fe-Mg silicate (Sil). The large central black region is a hole made by ion-thinning. The new phase, FOLAM, is observed at the positions indicated by white arrows. Along the right side of the spherule, fine-grained silicate grains (0.5–2 μm across) are intergrown with FOLAM.

The opaque spherules in the Allende chondrules and aggregates are commonly complex assemblages of troilite, pentlandite, magnetite and Ni-Fe metal in varying combinations and proportions¹. Silicates, chromite, heazlewoodite, and whitlockite are also finely disseminated in the opaque phases. Because of their intimately intergrown and fine-grained character, the mineralogical relationships are more complex than can be determined by conventional petrographic techniques.

The combined techniques of transmission electron microscopy (TEM) and scanning electron microscopy (SEM), with energy dispersive X-ray (EDS) and electron energy loss (EELS)

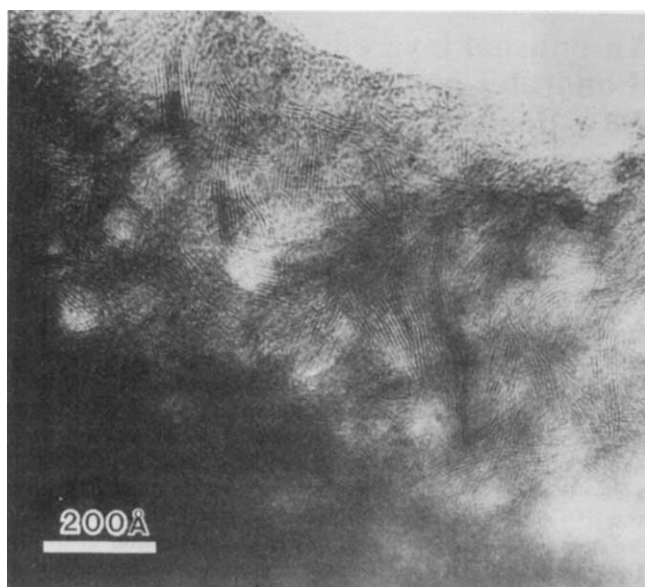


Fig. 2 A high-resolution TEM photo of FOLAM showing a complex arrangement of fringes, typical of materials having layer-structures.

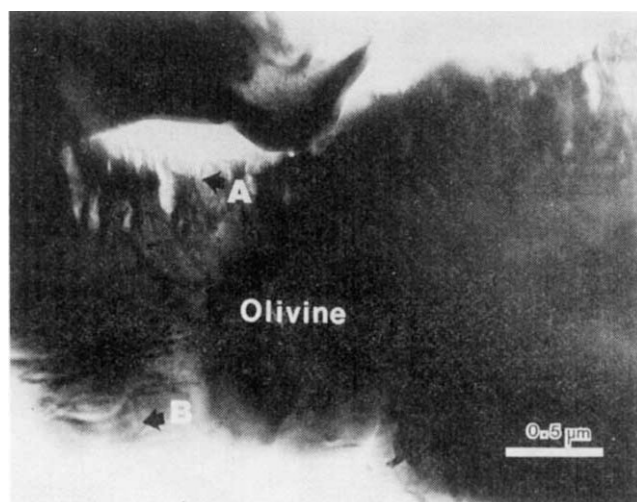


Fig. 3 A low-magnification TEM photo of an olivine grain showing the partial alteration to FOLAM at the top side of the grain. The grain occurs in an opaque spherule consisting of Ni-Fe, magnetite and pentlandite, within a barred olivine chondrule.

spectroscopy have been used to study the microstructures and phase relationships of the minerals in the opaque spherules. Chondrules and aggregates enclosing opaque spherules were extracted from a petrographic thin section and ion-thinned. The opaque minerals are softer than the surrounding silicate grains, and thus can be preferentially thinned for TEM observations.

Figure 1 shows an SEM photomicrograph of one of the ion-thinned opaque spherules. An unusual Fe-, Ni- and O-rich layered material (here called 'FOLAM' for brevity) is observed by TEM along the edges of the central hole. It commonly occurs in complex intergrowths with small silicate grains, as seen in Fig. 1. A low-magnification TEM image of FOLAM shows an extremely complicated, fluffy, 'cotton'-like texture. Each fragment appears to consist of a bundle of fine fibres entangled in complex ways. Submicrometre-sized chromite grains (<0.1 μm in diameter) commonly occur in association with FOLAM.

In most areas FOLAM is poorly crystalline and does not give clear diffraction patterns, but characteristically exhibits fairly strong powder rings at positions corresponding to spacings of 2.5, 2.0 and 1.5 $\text{\AA}$. Diffuse scattering characterizing spacings >2.5 $\text{\AA}$ suggests a wide variety of spacings in that range. Rarely, however, did we observe lattice fringe images (Fig. 2). Such images also exhibit extremely complex textures. The fringe spacings vary widely, but average $\sim 6.8 \text{\AA}$ which is close to but slightly smaller than the serpentine spacing (7.3 $\text{\AA}$). Most FOLAM crystallites consist of <10 layers and curve and diverge in various ways. The abundant overlapping and crossing fringes are especially prominent. The layers in these images are approximately parallel to the viewing direction; overlapping fringes therefore indicate that individual layers are not extended in the direction of the axes of each roll⁴. This unusual fringe texture can be best described as a 'ribbon'-structure resembling that in some graphitic carbons⁵. Based on the similarity of appearance in TEM images and spacings, we assume that FOLAM has a structure similar to that of serpentine.

EDS spectra from FOLAM show, without exception, large Fe and Ni and small Si, S, Cr and Mg peaks. Chlorine and Ca also occur, but in very small amounts. EELS spectra reveal a large oxygen peak in addition to the above elements. These compositional features differ from those observed in terrestrial phyllosilicates. Of special significance is the extremely small Si content, relative to the large amount of Fe, and the invariable presence of Ni, S and Cr. To explain the low Si content, we assume that FOLAM has a structure related to cronstedtite, an Fe-rich end member of the septeclorite series; cronstedtite is unique as a sheet silicate mineral, as it has Fe^{3+} substituting

for Si in tetrahedral sites⁶. It has been reported to occur in the Cochabamba C2(M) meteorite⁷.

As far as its chemical composition is concerned, FOLAM appears similar to 'PCP'⁸, a phase that is ubiquitous in C2 carbonaceous chondrites but has not yet been properly characterized. Previous petrographic studies suggest that PCP also has a layer structure and is closely related to phyllosilicates^{9,10}.

During our TEM study, we encountered evidence regarding the origin of FOLAM. Figure 3 shows an olivine grain that is partly altered to FOLAM along the top side. EDS analyses (Fig. 4) indicate that the olivine is distinctly more fayalitic than the olivines surrounding the opaque spherule. On the other hand, Si and Mg are strongly depleted while Fe and Ni are greatly enriched in the FOLAM. These observations show that FOLAM has been produced by alteration of olivines that are finely disseminated among the opaque phases.

Phyllosilicates are known to be major constituents of C1 and C2 carbonaceous chondrites, but relatively little is known about their occurrence and character in C3 carbonaceous chondrites. The matrix in Allende consists mostly of fine-grained olivines, and phyllosilicate has not been reported previously. It is significant that the FOLAM is restricted to the opaque spherules in chondrules and aggregates. Its occurrence suggests that conditions were once favourable for the formation of a serpentine-like mineral within the opaque spherules; presumably, low temperature and a somewhat hydrous condition were required. It seems improbable that the whole meteorite experienced such an environment, because otherwise fine-grained olivines, abundant in the matrix, would also have been affected and altered to sheet silicates.

Armstrong *et al.*¹¹ based on a detailed study of a hibonite-bearing inclusion in the Murchison C2(M) meteorite, showed

that aqueous alteration occurred before that inclusion was incorporated into its present location. Our observations provide an independent set of data that suggest an alteration stage may also have occurred in Allende before the final stages of meteorite formation. Either the environment required to produce FOLAM was unique to the opaque spherules, or the conditions needed to produce FOLAM occurred before the time that chondrules were trapped within the Allende matrix.

Haggerty and McMahon¹ argued that the presence of the unusually high Ni-content alloy coexisting with magnetite resulted from a highly oxidizing condition. The unusual composition of FOLAM presumably results from the oxidizing environment that the opaque spherules experienced during or after the formation of FOLAM. In these conditions, Fe would exist in the ferric state and thus be able to substitute for Si in the tetrahedral as well as octahedral sites of the cronstedtite structure. Ni²⁺ and Cr³⁺ must also enter the octahedral sites. Sulphur is conceivably incorporated into the structure as a sulphate ion. The occurrence of the large amount of Fe³⁺ could account for the smaller basal fringe spacing of FOLAM than those of the serpentine minerals. However, we are uncertain about the extent to which these ions can enter the structure without a breakdown of the lattice. In any case, the unusual characteristics of FOLAM reflect undetermined extraterrestrial conditions experienced by some chondrules and aggregates.

We thank John Larimer, Ian Mackinnon and David Veblen for helpful discussion and James Clark and Brian McIntyre for assistance with electron microprobe analyses and scanning electron microscopy. The microprobe was obtained partly with funds provided by the Chemistry Instrumentation Program of the NSF and the SEM is supported by NSF grant ATM 8022849. The electron microscopy was performed at the electron microscope facility in the Center for Solid State Science at Arizona State University. This work was supported by NASA grant NAGW-143.

Received 7 June; accepted 5 July 1982.

- McMahon, B. M. & Haggerty, S. E. *Proc. 11th Lunar planet. Sci. Conf.* 1003-1025 (1980).
- Haggerty, S. E. & McMahon, B. M. *Proc. 10th Lunar planet. Sci. Conf.* 851-870 (1979).
- McSween, H. Y. *Geochim. cosmochim. Acta* **41**, 1777-1790 (1977).
- Veblen, D. R. & Buseck, P. R. *Am. Miner.* **66**, 1107-1134 (1981).
- Ban, L. L., Crawford, D. & Marsh, H. J. *appl. Crystallogr.* **8**, 415-420 (1975).
- Hendricks, S. B. *Am. Miner.* **24**, 529-539 (1939).
- Müller, W. F., Kurat, G. & Kracher, A. *Tschermaks miner. petrogr. Mitt.* **26**, 293-304 (1979).
- Fuchs, L. H., Olsen, E. & Jensen, K. J. *Smithson. Contr. Earth Sci.* **10**, 1-39 (1973).
- Ramdohr, P. *The Opaque Minerals in Stony Meteorites* (Elsevier, Amsterdam, 1973).
- Bunch, T. E. & Chang, S. *Geochim. cosmochim. Acta* **44**, 1543-1577 (1980).
- Armstrong, J. T., Meeker, G. P., Huneke, J. C. & Wasserburg, G. J. *Geochim. cosmochim. Acta* **46**, 575-595 (1982).

Measurement of Rayleigh-Taylor instability in a laser-accelerated target

A. J. Cole & J. D. Kilkenny

The Blackett Laboratory, Imperial College of Science and Technology, London SW7 2BZ, UK

P. T. Rumsby, R. G. Evans, C. J. Hooker
& M. H. Key

Rutherford Appleton Laboratory, Chilton, Didcot,
Oxfordshire OX11 0QX, UK

The classical Rayleigh-Taylor (R-T) instability of a fluid supporting a higher-density fluid^{1,2} in a gravitational field is expected to occur in laser-driven compression³. A laser produces a high-pressure plasma that accelerates a higher-density solid shell of thickness Δr at a rate a . The effective gravitational acceleration— a would cause shell perturbations of wavenumber k to grow at the R-T rate $\gamma = (ka)^{1/2}$. However, ablation effects⁴⁻⁵ are predicted to damp short-wavelength modes, giving rise to a maximum growth rate at $k \sim 1/\Delta r$ and implying a maximum aspect ratio $r/\Delta r$ for a stable implosion. We present here the first clear experimental evidence for the growth of the

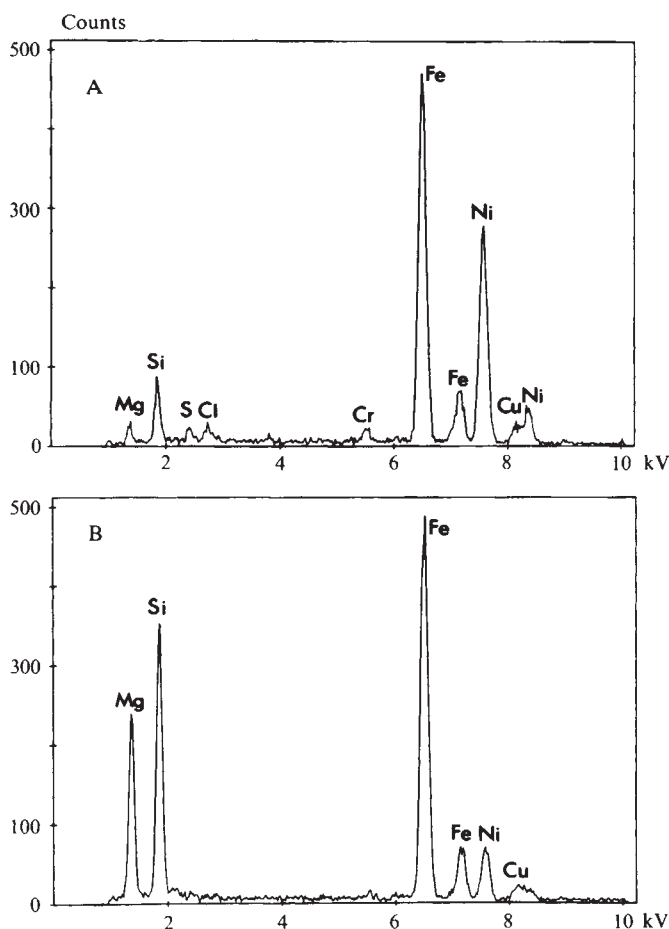


Fig. 4 EDS spectra from the areas of altered FOLAM (A) and olivine (B) in Fig. 3. The copper peaks result from the supporting grid.

R-T instability in a laser-accelerated plane target. These results were obtained by a novel technique using a target with an initial corrugation of known k . Streak X-ray radiography allows direct measurement of the growth of mass modulations in the target caused by this initial perturbation. The time-averaged growth rate of an imposed perturbation is measured as $(0.3 \pm 0.05) \times (ka)^{1/2}$, in close agreement with a numerical simulation.

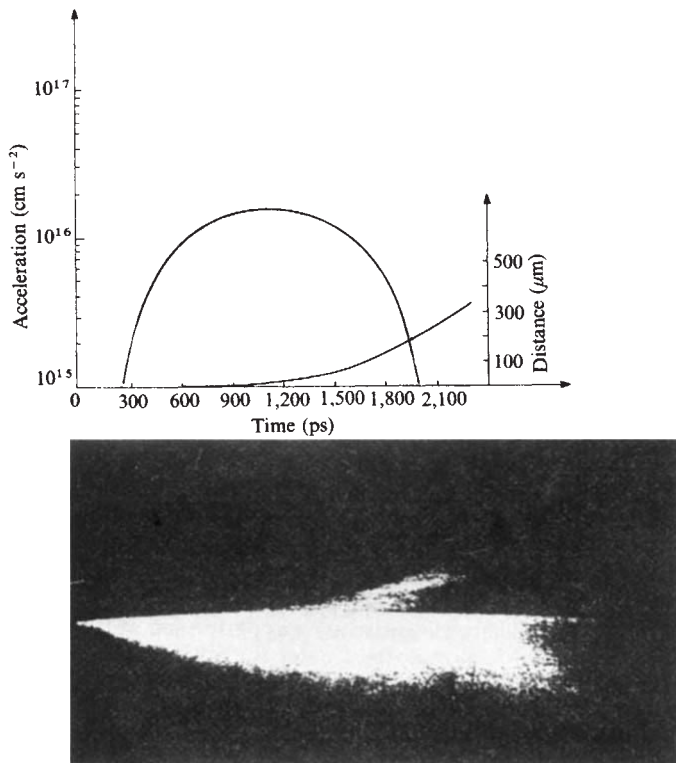


Fig. 1 X-ray streak camera photograph of a laser-accelerated microdisk target. The target is viewed edge on and is irradiated from below. The X-ray self emission of the accelerated material can be seen moving vertically away from the absorbing edge of unirradiated material. Also shown (above) is the target trajectory and the derived acceleration.

The accelerated target was a plane aluminium disk of diameter 550 μm and thickness 3 μm , symmetrically irradiated with three of the six orthogonal 0.53- μm beams of the Science and Engineering Research Council's Central Laser Facility⁶. Each of the beams was focused with an $f/1$ lens at an angle of incidence of 56°, allowing the target normal direction to be used for diagnosis by streaked X-ray radiography.

To measure their acceleration, the targets were positioned edge on and accelerated transversely across the field of view of an X-ray pinhole camera which had a 5- μm pinhole filtered with 3 μm of Al and 15 μm of Be. Images were recorded on an X-ray streak camera with 10 ps and 10 μm time and spatial resolution, respectively. The three beams with 32 J in total in 1.2 ns were focused 250 μm beyond the targets, thus only illuminating and accelerating the central 300 μm of the target. The streaked X-ray emission is shown in Fig. 1, where the accelerated central part of the target can be clearly seen moving away from the absorbing edge of the unaccelerated outer ring of the target. From knowledge of the laser irradiation-time profile and independently measured ablation pressure-irradiance scaling relationships⁷, it was possible to determine the acceleration of the target as shown in Fig. 1. This acceleration profile is compatible with the measured target trajectory also shown in Fig. 1 and could be scaled to match the irradiance of the transmission measurement described below.

To observe the R-T instability, the targets were turned so as to face away from the streak camera, and a different cluster of three beams was used to accelerate the target towards the camera. A more uniform illumination at approximately the same peak irradiance was used, by focusing 550 μm beyond the targets and increasing the laser energy to 87 J on target.

The R-T instability of the Al target was initiated by using a target with a corrugation of wavelength 20 μm and amplitude 0.5 μm . Behind the Al disk targets a separate Cu target was irradiated by a fourth synchronous backlighting laser beam. This produced a plasma 500 μm across and 100 μm high in the streak direction. In the spectral region of observation the X-ray emission from this target was several times brighter than that from the Al disk, as shown by an experiment with deliberately delayed backlighting. The Cu plasma provided a source with which to make an X-ray absorption picture of the accelerated Al disk. The spectrum of the Cu backlighting target, measured with two crystal spectrometers, showed a peak at a wavelength

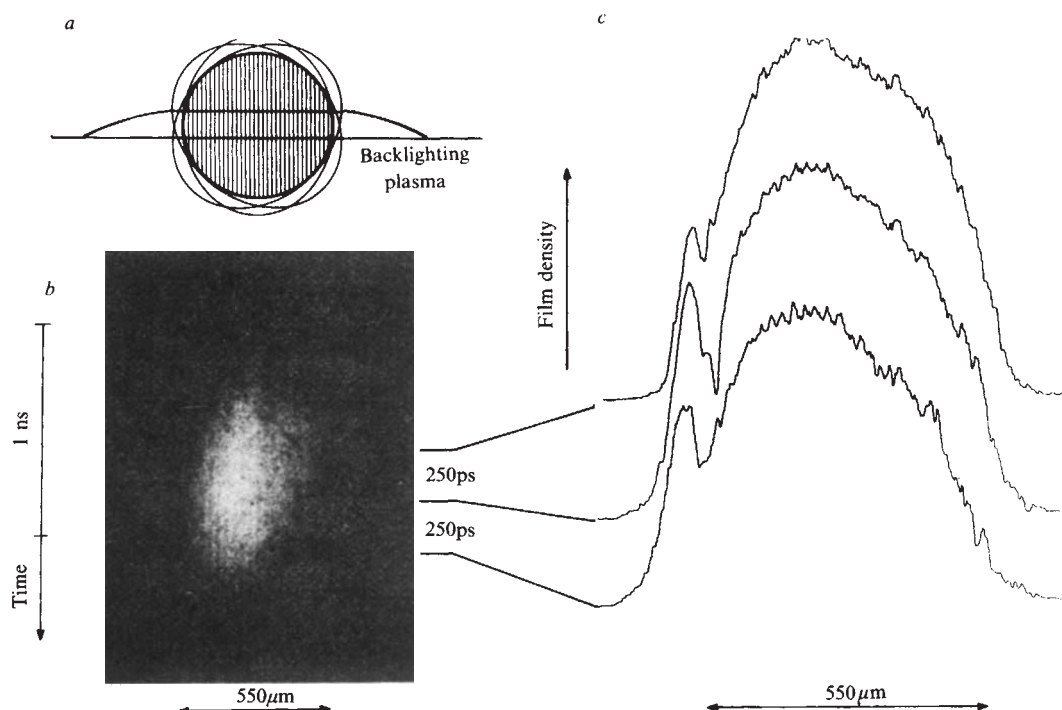


Fig. 2 *a*, Corrugated microdisk target illumination geometry showing location of backlighting plasma. *b*, X-ray streak photograph of backlighting emission through an irradiated microdisk target with 20 μm corrugation. *c*, Microdensitometer tracings of *b* at times 700, 950 and 1,200 ps, showing increasing modulation with time.

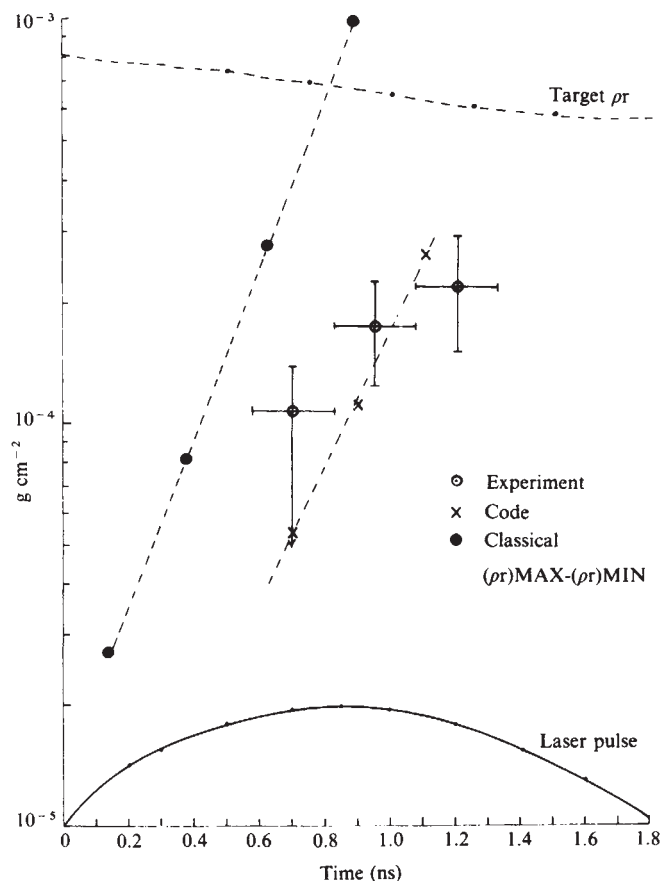


Fig. 3 Graph of modulation in target areal density from the three experimental measurements in Fig. 2c. Also shown are the results of two-dimensional fluid simulations and the modulation expected of a classically growing instability.

of $10 \text{ \AA}$ (in I_λ), with a rapid fall at shorter wavelength. The variation of the X-ray intensity reaching the streak camera with Al target thickness was calculated from the spectrum and the transmission of the filters, giving agreement with the 55% transmission observed through a $3\text{-}\mu\text{m}$ Al foil in a separate experiment.

An X-ray streak photograph of the transmission of such an accelerated target is shown in Fig. 2, with the corrugation parallel to the direction of streaking. Initially there is, as expected, no variation of transmission across the target, but as the instability grows, mass is redistributed and modulations in the transmission with a period of $20 \mu\text{m}$ can be seen. In control experiments with no initial corrugation, no such regular modulation is observed.

The film density modulations evident in Fig. 2 were converted to mass per unit area modulation by using the calculated variation of spectrally integrated transmission with Al thickness as described above. The amplitude of the modulation is shown in Fig. 3 as a function of time, the errors representing the spread in the amplitude of the eight modulations evident at late time in Fig. 2.

The average growth rate from Fig. 3 is $1.6 \pm 0.5 \times 10^9 \text{ s}^{-1}$, with some evidence of saturation. Figure 3 also shows the classical growth, taken from $(ka)^{1/2}$ and the acceleration scaled from the measurements shown in Fig. 1. The classical growth is started, in Fig. 3, at the onset of target motion with an initial amplitude determined by projecting back the slope of the experimental observations to zero time. The initial surface perturbation is not an R-T eigenmode, as there is no initial velocity perturbation, and this can cause some delay in the growth of an R-T mode⁸. However, it can be seen that the observed growth rate is $\sim 0.3 \times (ka)^{1/2}$.

The experiment has been simulated using a two-dimensional Eulerian fluid code⁹. An absorbed irradiance of $2.0 \times$

$10^{13} \text{ W cm}^{-2}$ was used so that the distance moved by the foil in the simulation matched the observed motion. A corrugated target as used in the experiment seeded the growth of an instability at the desired wavelength. The code predictions of growth in mass-density modulations (after convolving with a pinhole response) are also shown in Fig. 3.

The experimental measurements of Fig. 3 are lower than both the classical value and the simulated value. This is not due to ablating too much material from the target: the initial areal density is 0.8 mg cm^{-2} and the decrease due to ablation⁷ is shown in Fig. 3. The analysis of the Al transmission assumes the cold X-ray mass attenuation coefficients. However, the code values for the temperature and a local thermodynamic equilibrium ionization equation shows that this produces a negligible error. However, several phenomena are omitted or modelled inaccurately in the simulation, including radiation transport, equation of state, ion viscosity and magnetic field generation.

In conclusion, observations of the growth of area density modulations at a rate of $0.3 \times (ka)^{1/2}$ have been observed. Although these results are incomplete, in that data have so far only been obtained for one k value, it is nevertheless encouraging for laser-driven compression that the growth rate is so low, significantly less than classical, as has been predicted by numerical simulation⁵.

We acknowledge helpful conversation with W. C. Mead. A.J.C. was supported by a SERC studentship.

Received 13 May; accepted 14 July 1982.

1. Lord Rayleigh, *Theory of Sound* 2nd edn, Vol. 2 (Dover, New York, 1945).
2. Taylor, G. I. *Proc. R. Soc. A* **201**, 192 (1930).
3. Nuckolls, J. H., Wood, L., Thiessen, A. & Zimmerman, G. *Nature* **239**, 139 (1972).
4. McCrory, R. L., Monteith, L., Morse, R. L. & Verdon, C. P. *Phys. Rev. Lett.* **46**, 336 (1981).
5. Emery, M. H., Gardner, J. H. & Boris, J. P. *Phys. Rev. Lett.* **48**, 677 (1982).
6. Ross, I. *et al. IEEE J. Quant. Elect.* **QE17**, 1653 (1981).
7. Goldsack, T. J. *et al. Opt. Commun.* **42**, 55 (1982).
8. Lindl, J. D. & Mead, W. C. *Phys. Rev. Lett.* **34**, 1273 (1975).
9. Pert, G. J. *comp. Phys.* **43**, 111 (1981).

Density-driven interstitial water motion in sediments

D. L. Musgrave & W. S. Reeceburg

Institute of Marine Science, University of Alaska, Fairbanks, Alaska 99701, USA

Sediment depth distributions and fluxes of dissolved chemical substances have been interpreted as being a result of reaction, diffusion, bioturbation and irrigation^{1,2}. However, several studies suggest that density-driven convection³ can alter the depth distribution and increase the fluxes of dissolved substances when density decreases below the sediment surface⁴⁻⁷. We present here temperature-time series measurements for a freshwater lake undergoing autumn cooling. These are the first *in situ* observations of heat transport due to motion of interstitial waters over periods of less than 1 hour. Density, calculated from temperature, decreases with depth at the time and place that this motion occurs.

Temperature was measured by inserting a multi-thermistor plastic probe into the sediments. The probe was made of 2.5-cm PVC pipe with glass bead thermistors mounted along its length. The thermistors were pushed through drilled 2-mm diameter holes until the sensing element protruded at least 1 mm beyond the exterior wall of the pipe. Electrical leads from each thermistor were fed through the interior of the pipe to resistance bridges at the top of the probe. The probe was filled with polyurethane foam to reduce its thermal mass. Self-heating of the thermistors was negligible. Time response of the probe to temperature changes was $< 30 \text{ s}$. The voltage output from each thermistor was recorded on magnetic tape at 5-min intervals with a 12-bit Datal DL-2 data logger. All thermistors were corrected for

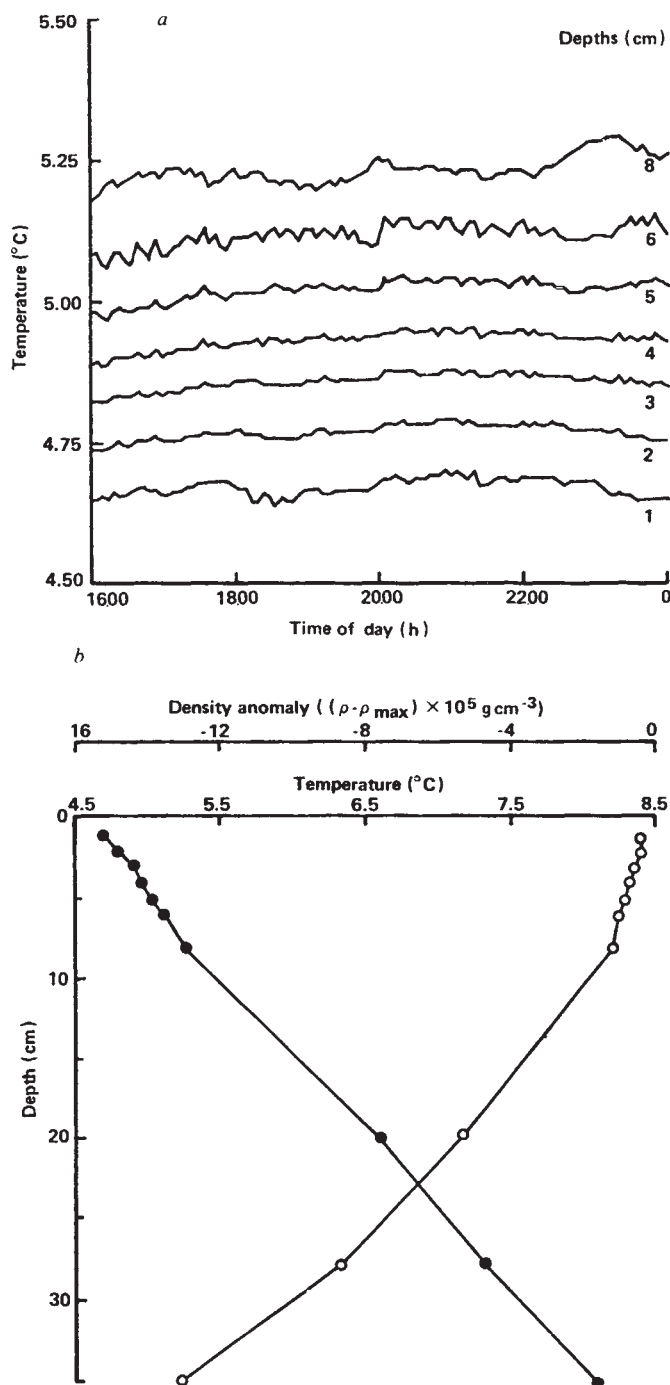


Fig. 1 *a*, Temperature records at depths below the sediment surface (27 October 1979; 0.75 m water depth). The record at 20 cm showed no temperature fluctuations. *b*, Temperature (●) and density (○) profiles in sediments (27 October 2000 CST). Density was calculated from temperature⁹.

gain and supply voltage changes by monitoring a reference voltage and the bridge supply voltage. The probes were calibrated at 2 °C intervals between 0 and 20 °C in a temperature bath controlled to 0.001 °C (Northwest Regional Calibration Centre). Although the probes were not recalibrated immediately after deployment, comparisons between calibrations indicated long-term (6-month) accuracy for each thermistor bridge of 0.01 °C. Short-term (1-day) stability is determined by the resolution of the data logger (1 bit in 12 over 27 ° = 0.007 °C), so changes in temperature can be resolved to this limit. The *in situ* temperature distributions were disturbed by the initial insertion of the probe for 4 hours. All temperatures reported here were from periods several days after the probe insertion.

We placed three probes in the sediments of Lake 227 of the Experimental Lakes Area, Ontario, Canada, at water depths of 0.75, 1.25, and 1.75 m. These sediments have high, nearly constant porosity (0.85) and organic content (48% dry weight measured by loss of weight on ignition at 900 °C)^{4,8}. The particle size data for two similar, nearby lakes at ELA show the following distribution⁸: sand (> 50 µm) 13%; silt (2–50 µm) 75%; clay (< 2 µm) 12%. The probes were in place from mid- to late-October 1979, in anticipation of bottom-water density increases due to autumn cooling. The diurnal temperature range at the sediment surface of Lake 227 was < 1 °C. The temperature increased with depth into the sediments (Fig. 1*b*). Temperatures varied smoothly at all depths until 25 October. At this time the temperatures in the upper 8 cm of the sediments at 0.75 and 1.25 m water depth developed transient perturbations that were independent of temperature variations at the sediment surface (Fig. 1*a*). The perturbations continued (with maximum amplitudes of 0.06 °C at depths of 6 and 8 cm) until the probes were retrieved.

The observed temperature fluctuations are probably caused by interstitial water motion due to buoyancy forces. However, they may be caused by other mechanisms, namely: (1) noise generated by the data acquisition system; (2) interstitial water motion caused by biological processes such as bioturbation and irrigation; and (3) interstitial water motion due to physical forces other than buoyancy.

We rely on our own calibration data and the long-term behaviour of the data acquisition system (as indicated by the temperature–time series) to address questions of noise due to instabilities of the acquisition system. Any noise-like error due to electronic instabilities of the data logger should appear in all channels because all input signals are processed in serial mode and thus employ the same circuitry. Visual inspection of the time series shows that the temperature fluctuations occur simultaneously only over a few adjacent thermistors. The relative stability of thermistors at other depths indicates that the electronics were introducing no sources of noise (and that the probe was not being physically moved in to or out of the sediment). There are two facts that indicate that the fluctuations are not caused by failure of electronics outside of the data logger: (1) before 25 October, fluctuations in the frequency range of the observed fluctuations do not occur in any channel; (2) the fluctuations in adjacent thermistors are coherent. Therefore, we are quite confident that we are observing *in situ* temperature fluctuations.

The possibility that the fluctuations are caused by bulk sediment redistribution or irrigation by infauna is not consistent with this environment. Hesslein⁴ showed that the solid phase of these sediments has not been redistributed after deposition. Infauna, if present, would disturb the sediment solid phase and therefore we take Hesslein's remarks as evidence that no infauna exist in these sediments. The absence of temperature fluctuations before 25 October corroborates this conclusion. We rule out the possibility of bubble ebullition for the same reasons.

Physical mechanisms that could cause interstitial water motion include ground water seepage, surface wave phenomena and buoyancy. Ground water flow is negligible, which is consistent with absence of the fluctuations for long intervals of time. In any case, we do not expect fluctuations of ground water flow with the regularity or high frequency of the observed fluctuations. Forced convection in the sediments due to surface waves or seiches is precluded by the lack of coherent temperature fluctuations in the upper sediments and concomitant attenuation of the surface fluctuations with depth in sediments.

The temperature gradient is destabilizing below the sediment surface due to autumn cooling of the bottom waters (Fig. 1) and may cause density-driven convection, leading to the observed temperature fluctuations. Quay¹⁰ shows that dissolved solids can significantly affect water density in Lake 227. The main effect in this context would be an offset in the density

since all solute gradients are above depths of 10 cm. The density gradients would not be affected.

The possibility of density-driven (free) convection leads us to consider the Rayleigh number³, Ra , defined by

$$Ra = \frac{\alpha \Delta T K (\rho_c)_i H}{\lambda_B}$$

where K is the hydraulic conductivity, α is the thermal expansion coefficient, $(\rho_c)_i$ is the heat capacity of the interstitial water, λ_B is the bulk thermal conductivity coefficient of the saturated porous medium, and H and ΔT are the scale depth and temperature difference, respectively. This dimensionless number is a measure of the ratio of the destabilizing buoyancy force to the stabilizing process of diffusion. According to theory and laboratory experiments, the thermal Rayleigh value must exceed a critical value ($Ra_c = 40$) before the motions induced by small horizontal temperature perturbations can maintain free convection. This critical value is derived assuming a saturated, homogenous, porous medium of depth H bounded above and below by impermeable, isothermal plates across which a destabilizing temperature difference, ΔT , is imposed. The equation of state relating density to temperature is linear and the system is assumed to be time independent. Nonlinear density-temperature relationships¹¹, periodic boundary conditions¹², nonhomogeneous¹³ or anisotropic porous medium¹⁴ or permeable boundaries¹⁵ can all act to reduce the critical Rayleigh number, although probably not by a factor of 10. Therefore, we take the value of 40 as an order of magnitude estimate for the critical Rayleigh number.

The Rayleigh number was estimated using the following assumptions. The high porosity permits the use of the thermal diffusivity, κ , of water ($1 \times 10^{-3} \text{ cm}^2 \text{ s}^{-1}$) for the quotient $\lambda_B/(\rho_c)_i$. Our efforts at measuring *in situ* hydraulic conductivity failed in Lake 227 so we estimated it to be 10^{-1} cm , which is similar to the value for sands. This is a high value but we think it is justified by the flocculent nature of Lake 227 sediments. The amplitude of the annual temperature cycle at the 1 m water depth is 20°C which we take for ΔT and the sediment depth, H , is 10 m. Using these values the maximum value of the Rayleigh number is 200, which is realized only when the amplitude of the annual temperature cycle is realized across the whole depth of the sediments. When this happens the interstitial water is freely convecting.

However, there is a period of time after autumn cooling begins when ΔT is not realized over the whole sediment depth. In transient systems, a time-dependent Rayleigh number can be calculated based on the product $H\Delta T$ where H and ΔT are scaled according to diffusion theory and both change with time. When this time-dependent Rayleigh number first exceeds the critical value at time t_c convection ensues over the depth $H(t_c)$ (ref. 16). The convecting layer erodes the stable layer below and H increases faster than purely diffusive transport would allow until the convecting layer reaches the sediment basement and ΔT is realized across the entire sediment depth.

Since we do not have temperatures below 35 cm we estimate H and ΔT as a function of time as follows. The temperature of the bottom water in contact with the sediments at 1 to 2 m water depth was 22°C from 1 June to 1 September. For this time interval we calculate the diffusive scale length to be 200 cm, which is the depth to which summer heating occurs. We assume that on 1 September the sediments to 200 cm are at 22°C , and the sediment surface begins cooling linearly with time to the observed temperature of 4.4°C on 30 October. The temperature-depth distribution as a function of time, for a system undergoing surface cooling at a linear rate, is given by Carslaw and Jaeger¹⁷. If we define the scale depth, H , as the depth at which the temperature is 0.0567 times the temperature difference between the temperature much deeper in the sediment and the temperature at the surface, then H is given by $H = 2\sqrt{\kappa t}$ where t is the time after 1 September. The scale temperature difference, ΔT , is then given by $\Delta T = mt$ ($0.0567 - 1$), where m is the rate of cooling. The product $H\Delta T$ increases

as the $3/2$ power of time and on 28 October is estimated to be $2 \times 10^3^\circ\text{C cm}$.

The Rayleigh number on this date is 20 and certainly within the range of the critical value, considering that this estimate is valid over an order of magnitude. The appearance of the temperature fluctuations at this time indicates that the estimate is fairly good.

Although we think the temperature fluctuations are due to free convection, they are not characteristic of linear, steady-state, convection. Horne and O'Sullivan¹⁸ have shown that oscillatory convection occurs at Rayleigh numbers above 240. Our estimate of the Rayleigh number is lower by 10 times, although we think that it could be an underestimate. However, the absence of temperature fluctuations below 15 cm is not consistent with the oscillatory systems described by Horne and O'Sullivan.

The density calculated from temperature shows negligible gradients in the upper few centimetres due to the density extremum at 4°C . So at the time the fluctuations occur, the unstable convecting layer is bounded above by a stable quiescent layer. Free fluids near the density extremum at 4°C show large temperature fluctuations at the boundary between the stable and convecting layers¹⁹. These fluctuations have been interpreted as internal waves propagating along the interface²⁰. The dynamics of flow in porous media do not allow inertial effects, and phenomena such as internal waves do not exist. However, we think that the convecting layer, beneath a shallow stable layer, causes the observed fluctuations in some manner whose details have been examined theoretically.

The increase in heat transport due to convection is quantified by the Nusselt number which is the ratio of heat flux across a convecting layer to the heat flux across the same layer assuming purely diffusive transport. The Nusselt number does not strictly apply to transient convection but we can get an order of magnitude of the increased transport from theoretical and experimental studies which give the Nusselt number as a function of the Rayleigh number³. Near the critical Rayleigh number the Nusselt number is just greater than 1. In Lake 227 the Rayleigh number continues to grow with time until the convecting layer intersects the impermeable rock below the sediments at 10 m. The Rayleigh number based on this depth is 200 and the maximum Nusselt number is then 4. The increase in transport of solutes cannot be readily estimated. In systems that are undergoing circular motion, there is no net water transport across any horizontal surface and net transport of solutes occurs only if vertical gradients in concentration exist. This implies that the increase in solute transport will also depend on the strength of the solute sink (or source) in the sediments. We know of no theory to describe the increase in transport of a passive solute as a function of the Rayleigh number and sink strength. We only mention that solute diffusivities are typically 10^{-2} – 10^{-3} times the thermal diffusivity and therefore seem more susceptible than heat to increased transport by convection.

This work was supported by NSF grants OCE 78-20896, OCE 80-26046 and a Jessie Smith Noyes Foundation Fellowship to D.L.M. We thank R. H. Hesslein and D. W. Schindler for their assistance, and W. D. Harrison for helpful discussions. Contribution no. 489 from the Institute of Marine Science, University of Alaska.

Received 21 December 1981; accepted 9 July 1982.

1. Berner, R. A. *Principles of Chemical Sedimentology* (Princeton University Press, 1971).
2. Berner, R. S. *Early Diagenesis. A Theoretical Approach* (Princeton University Press, 1980).
3. Combarous, M. A. & Bories, S. A. *Adv. Hydrosci.* **10**, 232–309 (1974).
4. Hesslein, R. H. *Can. J. Fish. aquat. Sci.* **37**, 545–551 (1980).
5. Thorstensen, D. C. & Mackenzie, F. T. *Geochim. cosmochim. Acta* **28**, 1–19 (1974).
6. Harrison, W. D. & Osterkamp, T. E. *J. geophys. Res.* **83**, 4707–4712 (1978).
7. Smetacek, V. F., von Bedungen, F., von Brockel, K. & Zeitzschel, B. *Mar. Biol.* **34**, 373–378 (1976).
8. Brunsell, G. J., Povedo, D., Graham, B. W. & Stainton, M. P. *J. Fish. Res. Bd Can.* **28**, 277–294 (1971).
9. Kell, G. S. *J. Chem. Engng Data* **12**, 66–69 (1967).
10. Quay, P. D. thesis, Columbia Univ. (1977).
11. Ray-shing Wu, Cheng, K. C., & Craggs, A. *Numerical Heat Transfer* **2**, 303–318 (1979).
12. Chhuon, B. & Calitigone, J. P. *J. Heat Transfer* **101**, 244–248 (1979).
13. Green, T. & Freehill, R. L. *J. appl. Phys.* **40**, 1759–1762 (1969).

14. Wooding, R. A. *Proc. 2nd Workshop Geotherm Reservoir Engng*, 339–345 (Stanford University Press, 1976).
15. Nield, D. A. *Wat. Resour. Res.* **4**, 553–560 (1968).
16. Elder, J. W. *J. Fluid Mech.* **32**, 69–96 (1968).
17. Carslaw, H. S. & Jaeger, J. C. *Conduction of Heat in Solids* 2nd edn (Oxford University Press, 1959).
18. Horne, R. N. & O'Sullivan, M. J. *J. Fluid Mech.* **66**, 339–352 (1974).
19. Townsend, A. A. *Q. Jl R. met. Soc.* **90**, 248–259 (1964).
20. Townsend, A. A., *J. Fluid Mech.* **24**, 307–319 (1966).

Thermal expansion effects in deep-sea sediments

T. J. G. Francis

Institute of Oceanographic Sciences, Brook Road, Wormley, Godalming, Surrey GU8 5UB, UK

Understanding how *in situ* heating affects un lithified deep-sea sediments is important for both geological and practical reasons. Deep-sea drilling by *Glomar Challenger* over the past decade has revealed several places where magma has intruded sediment to form sills and the heat released by a large sheet of molten rock has been shown to produce substantial changes on the sediments in its vicinity¹. Another example is the burial of hot canisters of high-level radioactive waste within the sediments of the ocean floor, which has been proposed as a solution to the problem of disposing of this material². I consider here the physical effects which high temperatures might produce in deep-sea sediments. Substantial chemical effects are also likely, but these are not discussed. In practice, the physical and chemical effects would work together in modifying the original sediment.

The individual thermal behaviour of the components of a sediment is straightforward and well understood. The pressure–volume–temperature relations for seawater differ very little from those for pure water³. The specific volume of pure water as a function of temperature and pressure has been tabulated by Burnham *et al.*⁴ and can be corrected with the work of Saunders⁵ to give values for seawater. The resulting thermal expansion of seawater at pressures of 200 and 500 bar is shown in Fig. 1. It is clear that the volume thermal expansion coefficient, defined as $1/V(\partial V/\partial T)_P$ is not constant but increases with increasing temperature. In contrast, the thermal expansion of most rock-forming minerals can be adequately described by a constant coefficient in the temperature range 0–400 °C (ref. 6). The thermal expansion of one particular solid material, quartz, is also shown in Fig. 1 to illustrate the contrast in thermal expansivity between the solid and liquid components of the sediment. One can conclude that the thermal expansivity

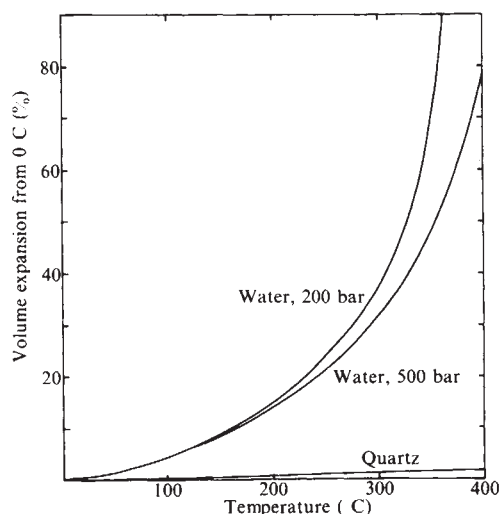


Fig. 1 The thermal expansion of water/seawater at 200 and 500 bar. The expansion of quartz, measured at atmospheric pressure, is shown for comparison. As a general rule the thermal expansivity of sedimentary pore water is two orders of magnitude greater than that of the solid component.

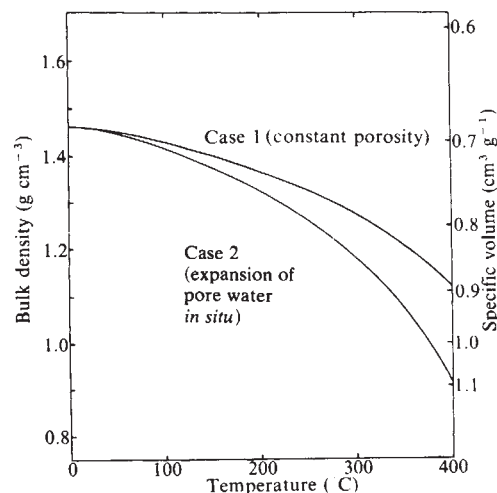


Fig. 2 The effect of thermal expansion on the density of a sediment at an ambient pressure of 500 bar. Initial porosity of sediment 0.75; grain density 2.7 g cm^{-3} .

of sedimentary pore water is some two orders of magnitude greater than that of the solid component. For the present purposes, the expansion of the solid component of the sediment is regarded as negligible.

The manner in which a sediment expands will be affected by the way in which heat is applied. The heating of sediment *in situ* is unlikely to be a uniform process. With magma intrusion and the burial of radioactive waste strong temperature gradients are likely to exist, both during the development of the temperature field and in the steady state. Two limiting cases to the way in which a sediment responds to heating can be considered. The first is the fully drained case, where the sediment matrix remains unchanged and expanding pore water is accommodated by flowing away from the source of heat. The porosity of the sediment is unchanged, but the bulk density of the heated sediment is reduced because of the reduced density of the pore water. The bulk density ρ is related to the solid particle density ρ_s and the fluid density ρ_f by $\rho = (1 - \phi)\rho_s + \phi\rho_f$ where ϕ is the fractional porosity. This expression can be used to calculate the density of the heated sediment since ρ_f is known as a function of temperature and ρ_s and ϕ are unchanged (case 1, Fig. 2). The second limiting case is where the pore water remains in place as it expands, disrupting the sedimentary matrix in the process. The bulk density of the heated sediment will be reduced both because of its increased porosity and because the density of the pore water is reduced. If a volume of sediment V weighs W and the weights of the solid and fluid components are W_s and W_f respectively, then

$$V = \frac{W_s}{\rho_s} + \frac{W_f}{\rho_f} = \frac{W}{\rho}$$

where in the undrained case W , W_s , W_f and ρ_s are constants. For a sediment of initial porosity 0.75 and grain density 2.7 g cm^{-3} , the expression becomes

$$\rho = \frac{1.4633}{0.25 + (0.7883)/(\rho_f)}$$

Values obtained from this equation corresponding to an ambient pressure of 500 bar are shown as case 2 in Fig. 2.

The behaviour of a sediment in reality will fall between these two limiting cases. One would expect a highly permeable sand to behave in a fully drained manner (case 1). A uniformly heated, impermeable clay might tend to follow case 2. As already mentioned, however, uniform heating is unlikely and strong temperature gradients are probable. In the second case, these would give rise to pore pressure gradients, pore water flow would occur to relieve the pressure and collapse of the expanded sediment matrix back towards its original porosity would result. Pore water flow will be facilitated by the increase

in permeability which accompanies an increase in porosity. Measurements of permeability as a function of porosity or void ratio have now been made on a range of deep-sea sediments (refs 7, 8, and P. J. Schultheiss, personal communication). These indicate that the increase in porosity associated with case 2 thermal expansion would be accompanied by an increase in permeability of at least an order of magnitude. Thus, although case 2 behaviour may be the instantaneous or short-term response of a sediment to heating, case 1 is more likely to reflect its longer-term behaviour. Nevertheless, case 2 is worth considering to determine the maximum reduction in density that a sediment might undergo.

Consider now the cooling of a sediment. When a highly permeable, coarse-grained sediment cools, the process of expansion is likely to be reversed. Water will flow back to fill the space created by the pore water contracting and the overall effect of the thermal cycle will be nil. But for fine-grained, impermeable sediments reversibility seems less likely. Thermal contraction of pore water will reduce pore pressure which, if not compensated by the return flow, will lead to collapse of the sedimentary matrix. Any collapse of the matrix will reduce its permeability, making return flow of pore water more difficult and further collapse of the sediment more likely. Thus the overall effect of heating and cooling sediment will be to produce a more compacted, less permeable sediment—a process which can conveniently be called 'thermal tamping'. Fine-grained sediments can be regarded as acting like a non-return valve, opening to allow increased pore water pressure to dissipate itself but closing to prevent pore water flowing towards zones of reduced pore pressure.

If thermal tamping is 100% efficient, the reduction in porosity or void ratio can be calculated as a function of temperature. If V_s is the volume of the solid component and V_v the volume occupied by the fluid component of the original sediment, then the original void ratio is defined by $e_{\text{orig}} = V_v/V_s$. If heating the sediment to $T^\circ\text{C}$ from its ambient value causes the pore water to expand by 100%, then the volume V_v occupied by fluid at $T^\circ\text{C}$ becomes $V_v/(1+m)$ on cooling back to 0°C . Thus the final void ratio on completion of the thermal cycle is given by $e_{\text{final}} = e_{\text{orig}}/(1+m)$, or in terms of porosity $\phi_{\text{final}} = \phi_{\text{orig}}/(1+m(1-\phi_{\text{orig}}))$.

Values of m taken from Fig. 1 have been used to compute the reduction in void ratio which thermal tamping would bring about to a sediment with original void ratio 3.0 (porosity 75%) at an ambient pressure of 500 bar and are shown in Fig. 3. Clearly, the higher the temperature to which the sediment is heated the greater the reduction in void ratio. Heating to 350°C reduces void ratio by ~ 1 , corresponding to a decrease in

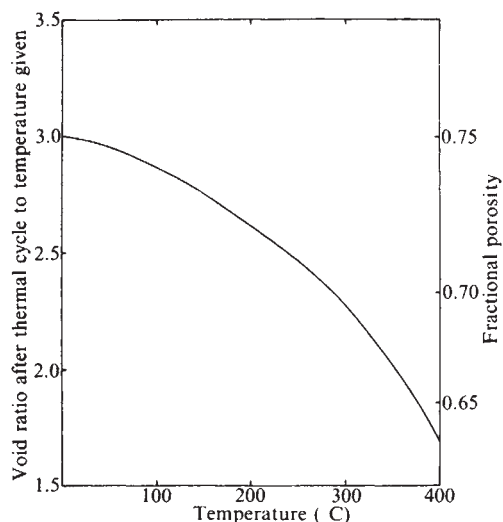


Fig. 3 Thermal tamping. The reduction in void ratio of a sediment, originally close to 0°C , when heated to a given temperature and then allowed to cool back to its original temperature. Original void ratio 3.0 (porosity 75%); ambient pressure 500 bar.

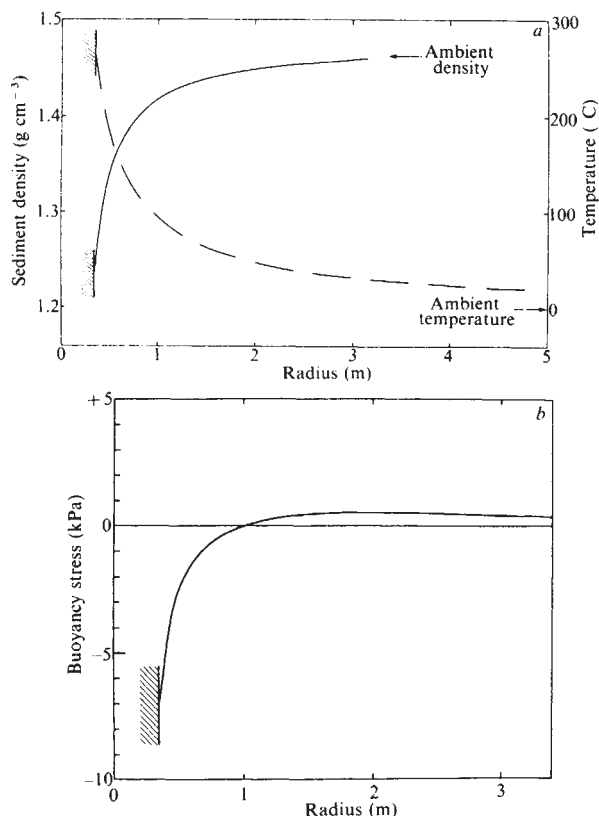


Fig. 4 *a*, The steady state temperature field around a spherical canister of radius 0.35 m, heat output $1,000\text{ W}$, embedded in sediment of thermal conductivity $0.85\text{ W m}^{-1}\text{ K}^{-1}$ (dashed line). The solid line shows the corresponding sediment density, assuming that the sediment expanded with the pore water *in situ* (case 2 in Fig. 2). *b*, The corresponding buoyancy stress of the canister and surrounding shell of sediment as a function of radius. Density of the canister, 3.0 g cm^{-3} . For comparison, the shear strength of a typical oceanic sediment at 30 m burial depth is $\sim 30\text{ kPa}$ (ref. 7).

permeability of approximately an order of magnitude.

Evidence for thermal tamping was obtained on Leg 64 of the Deep Sea Drilling Project, when *Glomar Challenger* drilled sites in the Gulf of California where basaltic sills had penetrated highly porous sediments at shallow depths⁹. The intrusion was found to reduce the porosity of the sediment in the vicinity of the sills¹. At hole 481A, for example, the original porosity of the sediment immediately above the sill is estimated at 0.7 and the final porosity is observed to be 0.3. The formula relating final to original porosity given above indicates that the pore fluid expanded by 444%. At an ambient pressure of 200 bar this requires that the temperature close to the sill reached $\sim 370^\circ\text{C}$ (ref. 1) (offscale on Fig. 1). Although the temperature of the molten rock must have been close to $1,200^\circ\text{C}$, a value in the region of 370°C is reasonable because this is about the temperature at which water boils at 200 bar. Furthermore, the relatively rapid contraction of the pore water in the region of 370°C may have aided the process of thermal tamping.

The ideas outlined above also allow us to calculate the physical properties of the sediment around a buried canister of radioactive waste. Immediately after emplacement the temperature field will develop, reaching an effectively steady-state configuration in a time of the order of a year¹⁰. Thereafter the temperature follows the decay of the radioactive waste contained in the canister. The peak temperature to which the canister wall rises must be restricted if a canister lifetime of 500–1,000 yr is to be achieved. At present, peak wall temperatures in the range $100\text{--}250^\circ\text{C}$ are thought appropriate¹¹. To understand the effect of thermally induced density changes in the sediment surrounding the canister, it is sufficient to rely on the peak temperature field developed around it. Making the following simplifying assumptions, this is easily calculated:

- (1) The canister is assumed to be a sphere of radius 0.35 m, making it comparable in volume with the more realistic cylindrical canisters proposed¹².
- (2) The heat output of the canister, Q , is taken to have a constant value of 1,000 W.
- (3) The thermal conductivity of the sediment, K , is taken as $0.85 \text{ W m}^{-1} \text{ K}^{-1}$ and assumed to be independent of the temperature. This is a good assumption up to 200°C and a reasonable one up to 400°C (ref. 12).
- (4) Hickox *et al.*¹⁰ have shown that thermally induced convection of the pore water is negligible and that conduction theory is sufficient to calculate the steady-state temperature field around the canister.
- (5) The ambient temperature of the sediment before emplacement is taken as 0°C .

With the above assumptions, the steady-state temperature of the sediment at radius r from the centre of the canister is given by

$$T = \frac{Q}{4\pi Kr}$$

For the values of Q and K given, this gives rise to the temperature distribution shown in Fig. 4a. Using the relationship of sediment density to temperature where the sediment expands with the pore water *in situ* (case 2 in Fig. 2), the corresponding density of the sediment as a function of radius has been obtained.

It has been suggested that heat from the canister would reduce the density of the sediment to such an extent that convective upwelling would occur which might carry the buried canister back to the sea bed (the 'Burp' effect). For this to be possible, the effective density of the canister and surrounding sediment must be less than the density of the unaffected sediment. The sediment density distribution shown in Fig. 4a allows this effect to be examined quantitatively.

Considering a spherical shell of heated sediment and outer radius r surrounding the canister of radius r_0 . Assuming that the canister remains fixed within the shell of heated sediment a buoyancy force, F , can be defined:

$$F = \frac{4}{3}\pi r^3 \rho_s - \frac{4}{3}\pi r_0^3 \rho_c - \int_{r_0}^r 4\pi r^2 \rho \, dr$$

where ρ_s is the density of unaffected sediment; ρ_c the mean density of canister; ρ the density of heated sediment $= \rho(r)$.

A positive buoyancy force will tend to move the canister upwards, a negative one to make it sink. To appreciate whether the buoyancy force developed is ever big enough to actually move the canister, it must be converted into a stress which can be compared with the strength of the sediment. A buoyancy stress can be obtained simply by dividing the buoyancy force by the cross-sectional area of the sphere: $S = F/\pi r^2$. S has been evaluated numerically taking values of ρ from Fig. 4a, $\rho_s = 1.463 \text{ g cm}^{-3}$ and $\rho_c = 3.0 \text{ g cm}^{-3}$ (see ref. 13) and is plotted in Fig. 4b. Beyond a radius of $\sim 1 \text{ m}$ the combined volume of canister and heated sediment is positively buoyant. The maximum buoyancy stress developed, however, is small, $\sim 0.6 \text{ kPa}$. In contrast the shear strength which the sediment can be expected to have at a burial depth of 30 m (ref. 13) is about 30 kPa (ref. 7). Because positive buoyancy is only developed beyond $\sim 1 \text{ m}$ radius and the maximum buoyancy stress at $\sim 2 \text{ m}$ radius, where the sediment is unlikely to be weakened either by temperature or by the mechanical effects of emplacement, positive buoyancy of the canister upwards seems very unlikely. This conclusion supports the results of the finite element modelling of Dawson and Chavez^{14,15}.

In conclusion, a hot body buried in deep-sea sediments will substantially modify the physical properties of the sediments adjacent to it. Cooling of the sediment is unlikely to be simply the reverse of heating, but could lead to collapse of the sediment matrix if pore water cannot flow back fast enough. This process, 'thermal tamping', would create a compacted shell of sediment around the body once it has cooled whose permeability would

be less than that of the unaffected sediment. The process could have important implications for the burial of hot canisters of radioactive waste beneath the ocean floor:

- (1) Thermal tamping would enhance the ability of the sediment to retain the radionuclides once the canister has corroded away. Since most of the thermal tamping would occur within 100 yr of burial, a canister life of $500\text{--}1,000 \text{ yr}$ may not be necessary.
- (2) The higher the peak temperature the more effective thermal tamping is likely to be. This would favour either a higher loading of radionuclides within the canister or a shorter period before disposal.
- (3) Thermal tamping would also be more effective at shallower sites, as the thermal expansivity of water decreases with pressure (Fig. 1). In this respect, therefore, a $4,000\text{-m}$ disposal site would be marginally more effective than a $6,000\text{-m}$ one.

An experimental investigation of the effects of temperature cycling on deep-sea sediments in *in situ* conditions is needed to confirm the existence and magnitude of thermal tamping.

I thank Peter Schultheiss and Hugh St John for helpful discussions. This research has been carried out under contract for the Department of the Environment, as part of its radioactive waste management research programme. The results will be used in the formulation of government policy, but at this stage they do not necessarily represent government policy.

Received 1 March; accepted 5 July 1982.

1. Einsele, G. *et al.* *Nature* **283**, 441–445 (1980).
2. Hollister, C. D. *Oceanus* **20**, 1, 18–25 (1977).
3. Bischoff, J. L. *Science* **207**, 1465–1469 (1980).
4. Burnham, C. W., Holloway, J. R. & Davis, N. F. *Geol. Soc. Am. spec. Pap. No. 132* (1969).
5. Saunders, P. M. J. *J. Phys. Oceanogr.* **11**, 573–574 (1981).
6. Skinner, B. J. *Geol. Soc. Am. Mem.* **97**, 75–96 (1966).
7. Silva, A. J. *Oceanus* **20**, 1, 31–40 (1977).
8. Silva, A. J., Laine, E. P., Lipkin, J., Heath, G. R. & Akers, S. A. *Proc. Marine Technology '80*, Washington DC (1980).
9. Curry, J. R. *et al.* *Geotimes* **24**, 7, 18–20 (1979).
10. Hickox, C. E., Gartling, D. K., McVey, D. F., Russo, A. J. & Nuttall, H. E. *Proc. Marine Technology '80*, Washington DC (1980).
11. Waste Form and Canister Task Group Report *Proc. 6th Annual NEA-Seabed Working Group Meet.* SAND-81-0427 (Sandia Laboratories, 1981).
12. Hadley, G. R., McVey, D. F. & Morin, R. *Proc. Marine Technology '80*, Washington DC (1980).
13. Systems Analysis Task Group Report *Proc. 6th Annual NEA-Seabed Working Group Meet.* SAND-81-0427 (Sandia Laboratories, 1981).
14. Dawson, P. R. *Buoyant Movements of Nuclear Waste Canisters in Marine Sediments*, SAND-78-0841 (Sandia Laboratories, 1978).
15. Chavez, P. F. & Dawson, P. R. *Thermally Induced Motion of Marine Sediments Resulting from Disposal of Radioactive Wastes*, SAND-80-1476 (Sandia Laboratories, 1981).

Composition of basaltic liquids generated from a partially depleted lherzolite at 9 kbar pressure

Gautam Sen*

Department of Geosciences, University of Texas at Dallas, Richardson, Texas 75080, USA

Two contrasting views exist regarding the nature and origin of mid-ocean ridge basalts (MORB): first, that these olivine-tholeiites with $\text{Mg}/(\text{Mg} + \text{Fe}) = 0.70\text{--}0.72$ are primary and are generated by partial fusion of plagioclase + spinel lherzolite at around 30-km ($\approx 9 \text{ kbar}$ pressure) in the mantle¹. Others^{2–8} suggest that even the most primitive MORBs are not primary but are derived by fractionation of olivine $\pm$ pyroxene from picritic parental magmas that are generated at depths of 45 km or more. Duncan and Green^{2,3} argued for the existence of a second primary magma, a high-MgO quartz-normative or olivine-poor tholeiite, which is produced at depths $\leq 25 \text{ km}$ beneath the ocean ridges. This problem, concerning the composition and depth of origin of primary MORBS, has been approached here by experimentally melting a plagioclase + spinel lherzolite xenolith from Hawaii at 9 kbar pressure. The experimental melts are very similar to some mid-ocean ridge basalts, which suggests that the latter are probably primary and are generated at shallow depths in the mantle.

* Present address: Department of Earth and Space Sciences, University of California at Los Angeles, Los Angeles, California 90024, USA.

The plagioclase + spinel lherzolite 77PAII-1 was collected from the Pali no. 2 vent⁹ on Oahu. The partially depleted character of this xenolith is reflected in its lower FeO, Na₂O and higher MgO contents relative to some of the least-depleted upper mantle peridotites (Table 1). Table 1 also shows that this xenolith is chemically similar to the plagioclase lherzolites of the Troodos ophiolite and the Trinity ultramafic complex. These bodies presumably represent mid-ocean ridge-type upper mantle^{2,10-16}, and therefore the results of the present study should be applicable to the problem of generation of mid-ocean ridge basalts.

The experiments were carried out in graphite capsules sealed within Pt-foil in a piston-cylinder apparatus. A piston-out technique was used, and no friction-correction was applied to the nominal pressure. The near-solidus and higher temperature runs were held for 26 and 6 h, respectively. The run products were analysed using an automated electron microprobe: 1–2 μ m diameter electron beam, 15 kV accelerating potential, and 0.15 μ A beam current. 10,000 or more counts were obtained for most major elements. Na-loss by volatilization was checked by simultaneously analysing three different Na-bearing standards. Quench crystals of pyroxenes were present in minor amounts; their abundance in the 1,400 °C and 1,600 °C runs made it impossible to obtain representative glass compositions for these runs. Fifteen or more spot analyses were obtained on glass (without quench crystals) from different parts of the capsule in each run. Although no reversal runs were made, general absence of zoning in the mineral phases, compositional homogeneity in the glasses, and overall correspondence between the analysed olivine and calculated equilibrium olivine compositions for the runs (Table 2) suggest that equilibrium was achieved in each of the runs. Results of a few runs performed with Ni and NiO starting materials, and the absence of metallic Fe in the run products suggest that the oxygen fugacity for these runs was between Ni/NiO and Fe/FeO buffers.

77PAII-1 begins to melt at a temperature of 1,210 °C, which is almost 100 °C lower than the solidus temperature in the CaO(C)–MgO(M)–Al₂O₃(A)–SiO₂(S) system¹ (Fig. 1). With increase of temperature, and degree of partial fusion, clinopyroxene (cpx) disappears at a temperature between 1,225 and 1,250 °C (Fig. 1, Table 1). Spinel (sp) continues to remain in the residue up to a temperature of ~1,290 °C. Olivine (ol) and orthopyroxene (opx) persist in the residue, until a temperature of ~1,600 °C where the residue becomes dunite, composed entirely of olivine (Fo₉₄).

Table 2 shows a comparison of compositions of the experimental melts with those of some proposed primary/primitive/parental magmas in mid-oceanic environments. B₁ and B₂ were postulated by Bryan¹⁷ to represent magmas allegedly

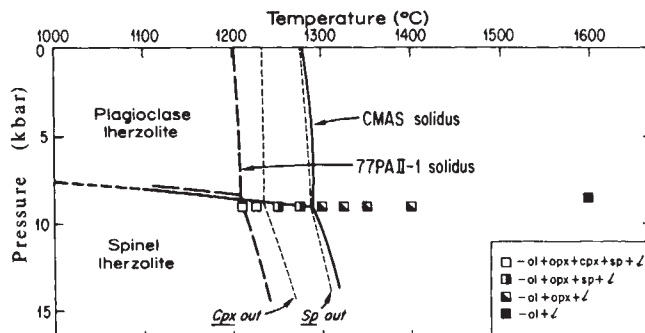


Fig. 1 Melting relations of the plagioclase + spinel lherzolite (77PAII-1) are compared with that of the CMAS system¹.

parental to the *Famous* (Mid-Atlantic ridge) basalt glasses. His "high-ol parental liquid" (B₁ in Table 2) appears to be a good candidate for a primary magma, following the definition of Presnall *et al.*¹. Note that B₁ is lower in SiO₂ and higher in Na₂O than the experimental melts, and this difference is manifested in the significantly higher content of normative olivine in B₁. Bryan's "low-ol parental liquid" (B₂ in Table 2) is much closer in composition to the melts produced at temperatures between 1,210 and 1,250 °C. B₂ is, however, higher in Na₂O than the experimental melts, which causes the former to be slightly olivine-normative. The "second-stage melts" of Duncan and Green², which were postulated to have been produced at a depth of ≤ 25 km, are also similar to B₂ and have more Na₂O than the experimental melts.

The difference in Na₂O between experimental melts of this study and the proposed primary/primitive/parental mid-ocean ridge magmas is not surprising in view of lower Na₂O content (that is, partially depleted composition) of the starting material than most relatively less-depleted peridotites (Table 1). Phase equilibria considerations of compositions and phenocryst assemblages of MORBs from different geographical localities¹⁸, major- and trace-element chemical studies, and isotopic studies^{19,20}, suggest that MORBs are likely to be derived from compositionally heterogeneous source lherzolites in the oceanic upper mantle. Noting that oceanic peridotites exhibit minor variations in FeO/MgO ratio but significant variations in Na₂O, CaO, Al₂O₃, and TiO₂, Bryan and Dick¹⁸ further argued that primary MORBs may show considerable spread in Na₂O, CaO, Al₂O₃, and TiO₂ contents but a very narrow range of FeO/MgO. Considering such major element heterogeneities in the ocean upper mantle and the Na₂O-poor composition of the starting material, the overall chemical similarity between the lower temperature experimental melts and the second-stage

Table 1 Compositions of the starting material and least depleted peridotites

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
SiO ₂	43.46	45.20	44.95	43.70	42.90	45.00	45.10	42.86	42.75	43.45
TiO ₂	0.07	0.71	0.08	0.25	0.20	0.07	0.50	0.33	0.02	0.04
Al ₂ O ₃	3.66	3.54	3.22	2.75	5.80	3.01	4.10	6.99	3.63	2.03
FeO*	7.60	8.52	7.67	10.19	9.20	7.98	9.90	9.33	8.25	8.23
MgO	42.42	37.50	40.03	37.22	37.20	39.70	36.70	35.07	40.34	43.78
CaO	2.00	3.08	2.99	3.26	3.70	3.15	2.30	4.37	3.46	1.67
MnO	0.12	0.14	0.13	0.14	0.11	0.11	0.20	0.14	0.18	0.14
Na ₂ O	0.02	0.57	0.18	0.33	0.40	0.24	0.60	0.45	0.06	0.06
K ₂ O	0.00	0.13	0.02	0.14	0.00	0.04	0.02	0.00	0.01	0.00
Cr ₂ O ₃	0.30	0.43	0.45	0.28	0.20	0.41	0.30	0.18	0.65	0.52
NiO	0.27	0.20	0.26	n.d.	0.20	0.25	0.20	0.20	0.25	0.24

(1) Plagioclase-spinel lherzolite 77PAII-1 (starting material used in this study); composition obtained from X-ray fluorescence.

(2) Pyrolite model composition⁵.

(3) Tinaquillo lherzolite⁵.

(4) PHN 1611, garnet lherzolite xenolith in South African kimberlite.

(5) Model upper mantle composition estimated by Carter²¹ for Kilbourne Hole.

(6)–(8) Analyses of lherzolites given by Ernst²².

(9) Average of three plagioclase lherzolites from the Troodos ophiolite¹¹.

(10) Average of six plagioclase lherzolite from the Trinity complex (source ref. 14).

Table 2 Comparisons of basalt compositions

Temperature (°C)§	9 kbar† melts (present study)							MORB ¹⁷		Second-stage melts ²			Plag-rich vein ¹⁶ 9W21
	1,210	1,225	1,250	1,275	1,300	1,325	1,350	B ₁	B ₂	(3)	(4)	(11)	
SiO ₂	50.61 (0.7)¶	51.64 (0.8)	51.44 (0.7)	51.65 (0.8)	51.18 (0.8)	51.34 (0.6)	51.12 (0.4)	48.53	51.20	54.20	52.40	53.30	49.31
TiO ₂	0.36 (0.06)	0.34 (0.08)	0.31 (0.06)	0.34 (0.08)	0.24 (0.07)	0.20 (0.07)	0.19 (0.04)	0.62	0.87	0.40	0.30	0.60	0.12
Al ₂ O ₃	17.73 (0.8)	15.53 (0.5)	15.08 (0.5)	15.06 (0.7)	13.98 (0.7)	13.50 (0.6)	11.25 (0.8)	16.65	15.03	15.40	11.70	14.40	15.59
FeO*	7.14 (0.3)	7.84 (0.5)	7.23 (0.4)	7.50 (0.5)	8.69 (0.4)	9.30 (0.6)	9.44 (0.8)	8.41	8.39	7.90	8.40	9.90	8.09
MgO	8.52 (0.5)	10.6 (0.3)	11.63 (0.7)	12.03 (0.6)	14.38 (0.4)	15.39 (0.5)	15.91 (0.4)	10.25	9.21	8.70	15.80	7.80	12.78
CaO	13.26 (0.6)	13.51 (0.4)	12.87 (0.8)	13.29 (0.4)	11.12 (0.5)	10.03 (0.5)	10.10 (0.6)	12.68	12.73	11.50	10.70	12.40	13.30
Na ₂ O	0.48 (0.09)	0.48 (0.05)	0.48 (0.04)	0.44 (0.04)	0.41 (0.05)	0.33 (0.07)	0.21 (0.03)	2.26	1.87	1.60	0.70	1.70	0.52
K ₂ O	0.10 (0.04)	0.07 (0.04)	0.09 (0.03)	0.05 (0.03)	0.05 (0.03)	0.07 (0.04)	0.07 (0.04)	0.06	0.10	0.10	0.10	0.10	0.02
Cr ₂ O ₃	0.03 (0.01)	0.03 (0.01)	0.02 (0.01)	0.29 (0.04)	0.36 (0.04)	0.47 (0.03)	0.48 (0.04)	0.03	0.07	n.d.	n.d.	n.d.	
CIPW norm													
q	4.92	3.28	3.11	2.28	0.36	0.14	0.97	—	—	5.47	—	3.59	—
or	0.60	0.41	0.54	0.29	0.29	0.41	1.42	0.25	0.57	0.59	0.59	0.59	0.12
ab	4.14	4.06	4.10	3.71	3.49	2.79	1.81	17.71	15.80	13.57	5.92	14.36	4.31
an	46.32	39.99	39.11	38.50	36.10	35.09	29.74	35.84	32.31	34.61	28.45	31.30	49.30
di	8.23	10.94	10.59	10.95	7.82	5.88	8.74	22.10	25.20	9.20	9.88	12.40	10.46
hy	26.54	29.73	29.25	32.41	43.96	49.32	49.34	—	21.65	26.61	44.65	24.31	27.66
ol	—	—	—	—	—	—	—	21.64	2.36	—	0.065	—	6.40
il	0.50	0.64	0.59	0.64	0.46	0.38	0.37	1.02	1.60	0.76	0.57	1.14	0.22
[Mg/(Mg+Fe)]*	0.887	0.902	0.903	0.909	0.910	0.914	0.916						
[Mg/(Mg+Fe)]†	0.876	0.889	0.905	0.905	0.908	0.909	0.910						

* [Mg/(Mg+Fe)] ratio in olivine analysed with the electron microprobe.

† [Mg/(Mg+Fe)] ratio in equilibrium olivine calculated on the basis of $K_D = (Fe/Mg)^{olivine} / (Fe/Mg)^{melt} = 0.3$ (ref. 23).‡ An uncertainty of ± 0.5 kbar exists.§ The nominal temperature values quoted are good within $\pm 10^\circ\text{C}$.

¶ Standard deviation given in parentheses.

melts is remarkable. Such compositional resemblance possibly suggests that at least some MORBs (second-stage melts) are primary and that they may be generated at shallow depths (~ 30 km).

A final comparison can be made with a proposed primary magma from the Trinity mafic-ultramafic complex¹⁶. These 'melts'¹⁶ occur as plagioclase-rich veins in host plagioclase lherzolite. Using compositional, field and phase equilibria arguments, Quick suggested that these melts were generated at pressures of <10 kbar. Table 2 shows that these veins are compositionally similar, including Na₂O contents, to some of the experimental melts, which supports the conclusion of Quick. It is further suggested that the source lherzolite for such melts must have undergone previous partial fusion and depletion in elements such as Na.

Application of the results of this experimental study to the MORBs and possible primary melts of the Trinity complex therefore supports the contention that the high-MgO, high-SiO₂ basalts of the mid-ocean ridges are primary, and are generated at relatively shallow mantle depths, possibly from plagioclase + spinel lherzolite.

This research was carried out at the University of Texas at Dallas. Financial support was received from NSF grants EAR74-22571A01, EAR8018359, and EAR7822766 to D. C. Presnall, and NSF grant EAR7923615 to W. G. Ernst. Reviews by Ernst and Presnall helped in revision of the original manuscript. Discussions with Presnall, Ernst, and J. E. Quick are gratefully acknowledged. G. Stummer provided the XRF analysis. C. McCann and V. Jones helped in preparation of the manuscript. This is Department of Geosciences, University of Texas at Dallas, contribution no. 427.

Received 19 April; accepted 29 June 1982.

- Presnall, D. C., Dixon, J. R., O'Donnell, T. H. & Dixon, S. A. *J. Petrol.* **20**, 3–35 (1979).
- Duncan, R. A. & Green, D. H. *Geology* **8**, 22–28 (1980).
- Duncan, R. A. & Green, D. H. *Geology* **8**, 562–563 (1980).
- Green, D. H., Hibberson, W. O. & Jaques, A. L. in *The Earth: its Origin, Structure, and Evolution* (ed. McElhinny, W. H.) 265–299 (Academic, New York, 1979).
- Jaques, A. L. & Green, D. H. *Contr. Miner. Petrol.* **73**, 287–310 (1980).
- O'Hara, M. J. *Nature* **220**, 683–686 (1968).
- Elthon, D. *Nature* **278**, 514–518 (1979).
- Stolper, E. *Contr. Miner. Petrol.* **74**, 13–28 (1980).
- Sen, G. thesis, Univ. Texas at Dallas (1981).
- Moore, E. M. & Vine, F. J. *Phil. Trans. R. Soc.* **268**, 443–466 (1971).
- Menzies, M. & Allen, C. *Contr. Miner. Petrol.* **45**, 197–213 (1974).
- Smewing, J. D., Simonian, K. O. & Gass, I. G. *Contr. Miner. Petrol.* **51**, 49–64 (1975).
- Smewing, J. D. & Potts, P. J. *Contr. Miner. Petrol.* **57**, 245–258 (1976).
- Quick, J. E. *Contr. Miner. Petrol.* **78**, 413–422 (1981).

- Lindsley-Griffin, N. *North American Ophiolites*, Bull. 95 (Department of Geology & Mineral Industries, Oregon 1977).
- Quick, J. E. *J. geophys. Res.* **86**, 11837–11863 (1981).
- Bryan, W. B. *J. Petrol.* **20**, 293–325 (1979).
- Bryan, W. B. & Dick, H. J. B. *Earth planet. Sci. Lett.* **58**, 15–26 (1982).
- Langmuir, C. H., & Hanson, G. *Phil. Trans. R. Soc. A* **297**, 383–407 (1980).
- Wilkinson, J. F. G. *Earth-Sci. Rev.* **18**, 1–57 (1982).
- Carter, J. L. *Bull. geol. Soc. Am.* **81**, 2021–2034 (1970).
- Ernst, W. G. *Petrologic Phase Equilibria* (Freeman, San Francisco, 1976).
- Roeder, P. L. & Emslie, R. F. *Contr. Miner. Petrol.* **29**, 275–289 (1970).

Deep structure of the Scottish Caledonides revealed by the MOIST reflection profile

David K. Smythe*, Alan Dobinson*, Robert McQuillin*, Jonathan A. Brewer†, Drummond H. Matthews‡, Derek J. Blundell‡ & Brian Kelk§

* Institute of Geological Sciences, Murchison House, West Mains Road, Edinburgh EH9 3LA, UK

† Department of Earth Sciences, University of Cambridge, Madingley Road, Cambridge CB3 0EZ, UK

‡ Department of Geology, Chelsea College, University of London, 552 Kings Road, London SW10 0UA, UK

§ Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wilts SN2 1EU, UK

The Moine and Outer Isles Seismic Traverse (MOIST) is a deep crustal reflection profile shot at sea off the north coast of Scotland in 1981. Spectacular reflections are observed from the Moho and from thrust zones within the Caledonian fold belt and foreland. Deep reflection profiling of the continental crust of the United States under the direction of the COCORP group, has amply demonstrated the value of this technique when applied to suitable geological problems (see, for example, refs 1, 2). A similar group, the British Institutes Reflection Profiling Syndicate (BIRPS), has now been set up in the UK to organize such projects. Funds are provided by the Natural Environment Research Council (NERC) through the core group based at Cambridge University. However, a preliminary experiment, of which MOIST is the result, was organized through the Institute of Geological Sciences (IGS).

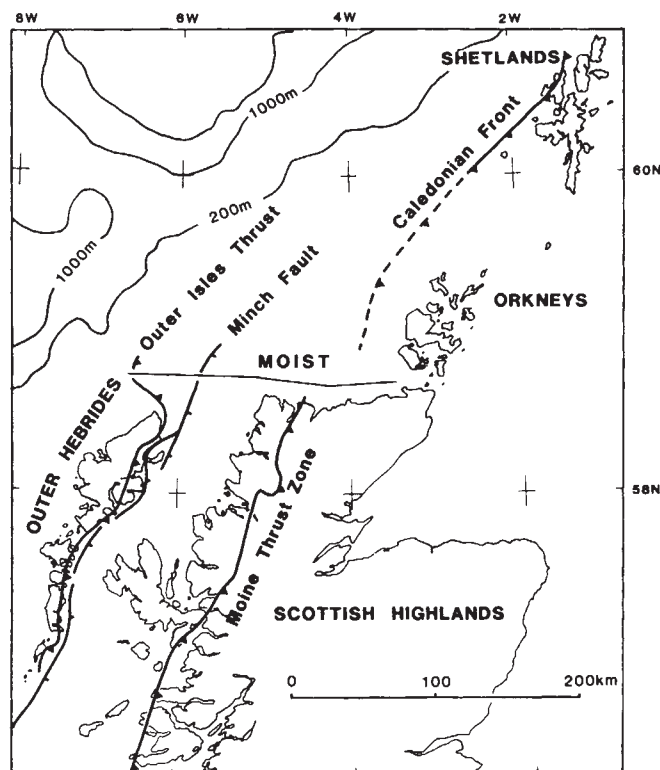


Fig. 1 Location of MOIST in relation to the Caledonian orogenic front off northern Scotland.

The strategy of recording reflection profiles offshore was adopted because a number of areas of basement outcrop around the United Kingdom are known to give rise to good upper- and mid-crustal reflections on confidential commercial reflection data³⁻⁵. Furthermore, shooting at sea is nearly an order of magnitude cheaper, mile for mile, than on land, as well as giving better quality data, although the obvious drawback is that reflections cannot be tied in directly to exposed outcrop.

One of the most important targets for reflection profiling in the UK is the Caledonian orogenic front (Fig. 1), which runs the length of north-west Scotland, trending north-east offshore from the Scottish mainland towards the Shetlands⁶. The nature of the front and the deeper crustal structure of the Moine Thrust zone and Outer Isles Thrust are the subject of several conflicting hypotheses, referred to below. There is a good coverage of industry reflection data offshore, recorded mainly to 5 s two-way time (TWT)—up to 15 km of crustal penetration—and several major crustal refraction profiles have been observed both offshore and onshore⁷⁻¹⁰. Although in many respects these long refraction profiles have proved disappointing in their lack of resolution of crustal problems, they do at least define well the Moho depth, which is about 25 km below the foreland.

MOIST crosses the offshore extrapolation of the Caledonian fold belt and runs onto the foreland to the west (Fig. 1). Its precise location was fixed with the help of the commercial data³, avoiding as far as possible the complicating effects of plutons and other intrusions. The survey comprised one line running from the Pentland Firth to 30 km NW of the Butt of Lewis, a total of 181 km (Fig. 1). The line consisted of three straight segments run close to and parallel to the coast for best possible correlation with known land geology, but sufficiently far offshore to minimize side swipe interference from charted coastal features.

The survey was conducted by Western Geophysical Company of America, using essentially the same equipment as for conventional oil industry surveys, except that the source and detector were towed at a greater depth to optimize the lower frequency signal content. The source was a 905 in³ tuned airgun array

operating at 4,500 p.s.i. pressure, towed at a depth of 12 m. The 3-km hydrophone streamer, comprising 60 × 50 m sections, was towed at a depth of 15 m. Shot interval was 50 m. Sixty channels were recorded at a 4-ms sampling rate to a record length of 15 s. Positioning was by satellite navigator integrated with bottom tracking Doppler sonar.

A conventional basic processing sequence was followed to produce final stacked sections with 30-fold subsurface coverage at a 25-m common depth point interval. In addition, array simulation techniques were used before stacking, to enhance the continuity of events deep in the section. The operator lengths were equivalent to a 200-m source array and a 100-m

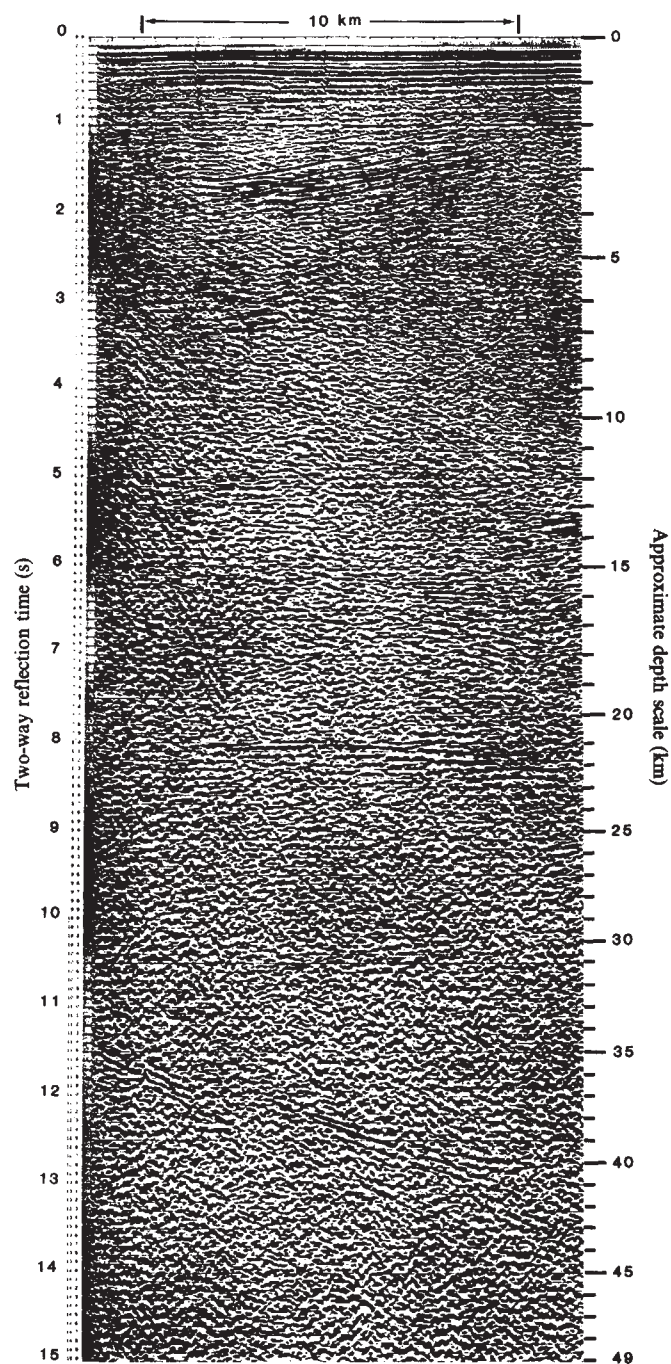


Fig. 2 Part of the final equalized stack section processed by Western Geophysical (located in Fig. 3). Strong reflectors at 1.4–1.8 s TWT are from Permo-Triassic redbeds below Jurassic shales. The Moho is at 8.0 s TWT and the 'Flannan Thrust' reflectors dip east between 12 and 13 s TWT. Depth scale on right is approximate.

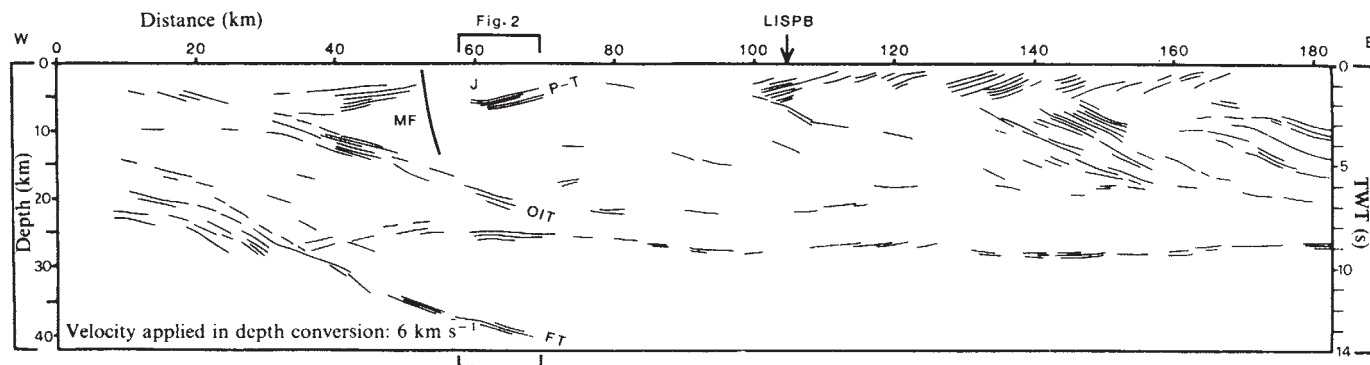


Fig. 3 Line drawing of MOIST profile, based on migrated sections but also using data from unmigrated stacks. Vertical depth scale assumes a constant velocity of 6 km s^{-1} ; no vertical exaggeration. LISP intersection near Cape Wrath is shown. MF, Minch Fault (drawn from other data); J, Jurassic; P-T, Permo-Triassic; OIT, Outer Isles Thrust; FT, Flannan Thrust.

receiver array. Velocity analyses were computed at 3-km intervals.

Figure 2 shows part of the equalized stack section resulting from the above processes, showing the excellent data quality. Figure 3 is a line drawing based on a migrated section produced for BIRPS by Shell Expro UK Ltd, but also using the information from the conventional stack (for example, Fig. 2) together with other migrations subsequently produced by one of us (J.A.B.) on the COCORP computer at Cornell University.

The base of the crust is well defined by an almost continuous Moho reflector along most of the section, 35–181 km, at a depth of around 25 km, in good agreement with the LISP and NASP profiles^{7,8}. Apparent vertical offsets in TWT, for example, astride the Minch fault, may be due to structural offsets in the Moho, although pull-up effects due to lateral variations in crustal P-wave velocity cannot be ruled out. The W-dipping reflectors in the uppermost 2–3 s TWT are from sediments in half-grabens of ages ranging from Proterozoic (?Stoer Group, 990 Myr) to Jurassic. Lewisian basement of Laxfordian (~1,700 Myr) or older age, forming the sub-sedimentary crop along the westernmost two-thirds of the profile, is the apparently autochthonous foreland. It generally lacks coherent reflectors, but in the western half of the profile is cut by two sequences of reflectors dipping east at around 20° . The upper of these corresponds to the northward offshore extension³ of the Outer Isles Thrust, a major intra-foreland thrust zone cropping out along the eastern coastline of the Outer Hebrides. The thrust appears either to flatten out at ~20 km depth or to cut and displace the Moho (Fig. 3, 70–90 km along profile). In either case, it has clearly been reactivated as a low-angle normal fault, permitting deposition of the westward dipping sediments above. Comparisons with the geologically better controlled geophysical data in the Minches and Sea of the Hebrides^{11–14} suggest that the Stoer Group (~990 Myr) and/or Torridon Group (~810 Myr) piedmont fan and fluvial deposits^{15–17} may form the base of these sedimentary prisms, in which case the Outer Isles Thrust would be of Grenville or earlier age.

The deeper sequence of eastward-dipping reflectors was totally unexpected. It appears to truncate the Moho (Fig. 3, 35 km along profile) and can be traced convincingly to 13 s TWT (Figs 2, 3)—about 40 km depth. The reflectors are primary, not multiple, events; their apparent termination of the Moho, and their parallelism with the overlying Outer Isles Thrust, suggest that the reflector group lies approximately in the plane of section, and that it is due to another thrust. We call it the Flannan Thrust, since it may crop out along the western margin of the Flannan Trough^{18–20} west of the Hebrides.

The west-dipping half-grabens seen along the eastern part of the section (Fig. 3, 100–170 km along profile) contain red beds of late Palaeozoic age. The easterly-downthrowing faults bounding the sedimentary prisms presumably flatten out into

the group of prominent east-dipping reflectors beneath, which have been interpreted³ as due to intra-orogenic belt structure. Our new interpretation of these requires a fuller discussion (J.A.B. and D.K.S., in preparation) than can be given here; however, the data suggest that thrust structures flatten out at 17–20 km depth. Thus, thin-skinned models^{21–23} of the Moine Thrust Zone and 'duplex' models of full crustal thickness²⁴ appear to be incorrect.

In conclusion, MOIST demonstrates that conventional marine seismic reflection techniques are indeed now capable of profiling the whole continental crust, contrary to pessimistic theoretical predictions²⁵, and also, that reflection profiling, even when located offshore, can be an extremely powerful tool for testing *a priori* hypotheses. We stress, however, that line drawings such as that in Fig. 3 cannot simply be 'coloured in' with a geological interpretation; new and better-constrained hypotheses can only emerge from a synthesis of all available geological and geophysical data.

We thank Western Geophysical, especially David Skidmore, for providing us with help and advice above their contractual obligations. We also thank Shell Expro UK Ltd, and particularly Gordon Robson and Rodney Calvert, for giving us both the benefit of their experience and for supplying us with a migrated version of the line.

Received 13 July; accepted 4 August 1982.

1. Cook, F. A. *et al. Geology* **7**, 563–567 (1979).
2. Ando, C. J., Cool, F. A., Oliver, J. E., Brown, L. D. & Kaufman, S. *Mem. geol. Soc. Am.* (in the press).
3. Smythe, D. K. *Geophys. J. R. astr. Soc.* **61**, 199 (1980).
4. Blundell, D. J. & Smythe, D. *Geophys. J. R. astr. Soc.* **61**, 216–217 (1980).
5. Blundell, D. J. in *Petroleum Geology of the Continental Shelf of North-west Europe* (eds Illing, L. V. & Hobson, G. D.) 58–64 (Heyden, London, 1981).
6. Flinn, D., Frank, P. L., Brook, M. & Pringle, I. R. in *The Caledonides of the British Isles—Reviewed* (eds Harris, A. L., Holland, C. H. & Leake, B. E.) 109–115 (Scottish Academic, Edinburgh, 1979).
7. Smith, P. J. & Bott, M. H. P. *Geophys. J. R. astr. Soc.* **40**, 187–205 (1975).
8. Bamford, D., Nunn, K., Prodehl, C. & Jacob, B. *Geophys. J. R. astr. Soc.* **54**, 43–60 (1978).
9. Hall, J. J. *geol. Soc.* **135**, 555–563 (1978).
10. Bott, M. H. P., Armour, A. R., Himsworth, E. M., Murphy, T. & Wylie, G. *Tectonophysics* **59**, 217–231 (1979).
11. Smythe, D. K., Sowerbutts, W. T. C., Bacon, M. & McQuillin, R. *Nature phys. Sci.* **236**, 87–89 (1972).
12. McQuillin, R. & Watson, J. *Nature phys. Sci.* **245**, 1–3 (1973).
13. Binns, P. E., McQuillin, R. & Kenolty, N. *Rep. Inst. Geol. Sci.* no. 73/14 (1974).
14. Smythe, D. K., Cheshier, J. & Bishop, P. *Rep. Inst. Geol. Sci.* (in the press).
15. Stewart, A. D. *Mem. Am. Ass. petrol. Geol.* **12**, 595–608 (1969).
16. Williams, G. E. *Mem. Am. Ass. petrol. Geol.* **12**, 609–629 (1969).
17. Stewart, A. D. *Spec. Rep. geol. Soc. Lond.* **6**, 43–51 (1975).
18. Jones, E. J. W. *Nature* **272**, 789–792 (1978).
19. Smythe, D. K., Kenolty, N. & Russell, M. J. *Nature* **276**, 420–421 (1978).
20. Jones, E. J. W. *J. geophys. Res.* **86**, 11, 533–511, 574 (1981).
21. Elliot, D. & Johnson, M. R. W. *Trans. R. Soc. Edinb. Earth Sci.* **71**, 69–96 (1980).
22. Coward, M. P. *J. struct. Geol.* **2**, 11–17 (1980).
23. McClay, K. R. & Coward, M. P. *Geol. Soc. Lond. Spec. Publ. No. 9*, 241–260 (1981).
24. Soper, N. J. & Barber, A. J. *J. geol. Soc.* **139**, 127–138 (1982).
25. Backus, M. M., Stoffa, P. L., Tsai, C. J. & Stark, T. J. *Geophysics* **47**, 454 (1982).

Oceanic plateaus as meteorite impact signatures

Garry C. Rogers

Pacific Geoscience Centre, Earth Physics Branch, Sidney, British Columbia, Canada V8L 4B2

The oceanic plateaus are an enigmatic set of deep ocean structures¹⁻⁵. Could these be signatures of ancient meteorite impacts? While numerous confirmed and suspected impact sites, ranging in age from Precambrian to Recent, have been identified on the continents⁶, none has thus far been identified in the oceans. The ocean floor is relatively young compared with most of the continental land masses and is constantly being renewed at spreading centres and destroyed in subduction zones. Nevertheless, roughly half of the ocean basins are of Cretaceous age or older (the Pacific ocean floor is shown in Fig. 1) and they represent geologically very stable platforms. Noting, for example, the number of major continental impacts documented to be less than 100 Myr old (15 craters with diameters >10 km including two >50 km in the Soviet Union⁶), a significant number of oceanic impact structures of this age should also be present. A recurrence relation derived from impact structures on land, when correlated for the size and average age of the ocean basins, indicates that impact structures should be more numerous in the oceans (Fig. 2). The key problem is what to look for. I discuss here reasons for considering the sparsely scattered deep ocean plateaus to be likely candidates.

Deep ocean plateaus can be divided into two categories⁴: continental type and oceanic type. The former seem to be rifted segments of continents and are not of concern here. The origin of the oceanic type of plateaus is much more difficult to explain. They are few in number and most are away from plate margins in older ocean floor. They are broad, high standing, aseismic features, with little relief in the crestal zone. The largest is several hundred kilometres in dimension. Their crustal structure differs markedly from the relatively uniform three-layer model which applies to most of the oceanic crust. They seem to be equivalent to thickened sections of oceanic crust³, with a notably thick basal layer having a compressional wave velocity in the 7.1–7.6 km s⁻¹ range⁴. Most plateaus do not exhibit significant isostatic anomalies, implying almost complete compensation⁵. The larger plateaus are capped by thick calcareous sediments, some of which are now below the calcite compensation depth, suggesting significant subsidence since formation. Sampling from the Deep Sea Drilling Project (DSDP) has confirmed a shallow water origin by the presence and size of abundant vesicles in the basalts cored and the presence of shallow water fauna in the fossil record⁷⁻⁹. The magnetic signature of the plateaus is confused and different from the lineations which characterize typical oceanic crust⁵, suggesting that normal seafloor spreading was not involved or that they were formed during one polarity interval of the Earth's magnetic field. Although recent authors agree that oceanic plateaus have been caused by unusually voluminous basaltic eruptions^{2,3}, no complete explanation for their origin has yet been published²⁻⁵.

With the properties of the plateaus in mind, suppose that a large extraterrestrial body (5–10 km in diameter) collided with the Earth in an oceanic area. The energy-absorbing effect of a few kilometres of water has been estimated to be negligible in a collision with a body of this size^{10,11}. The impact would be large enough to cause a crater of the order of 100 km in diameter¹⁰⁻¹¹ and fracture completely through the oceanic lithosphere, causing an initial ballistic structural uplift that would upwarp the asthenosphere, distort the geothermal

gradient and raise the deep ocean floor to near or above sea level. (A central uplift of 1/10th the crater diameter is not uncommon for large craters on land¹³⁻¹⁵.) If the impact was large enough, the upwarping of the asthenosphere might be sufficient to cause the formation of a long-lived thermal plume in the mantle¹⁰. In any case, massive outpouring of basalt and significant melting of the upper mantle might be expected if the bolide targeted on young ocean floor^{10,11}. (The plateaus of the western Pacific all have ages within a few tens of millions of years of the surrounding sea floor^{3,7,9,16}.) The volcanism would extend much beyond the original crater because of extensive fracturing of the thin oceanic lithosphere, and the original impact structure would thus probably be obliterated. The basalts erupted in shallow water on the top of the uplifted central structure would contain many vesicles, and the time constant for viscoelastic response of the lithosphere (of the order of 10⁶ or 10⁷ yr)^{17,18} would allow a large body of calcareous sediment to build up as it sank slowly back to equilibrium position. The mixture of water with the melted and uplifted mantle would cause widespread serpentinization, resulting in a net volume increase (and a thick, high velocity basal layer in the crust^{1,4}) which would make the final structure a plateau significantly above the abyssal sea floor. Thus, the properties of the oceanic plateaus are not inconsistent with what could be expected for structures caused by the impact of a large extraterrestrial body.

Consistency, however, is not proof. The diagnostic features of astroblemes on land^{13,19}—large circular structure, meteorite

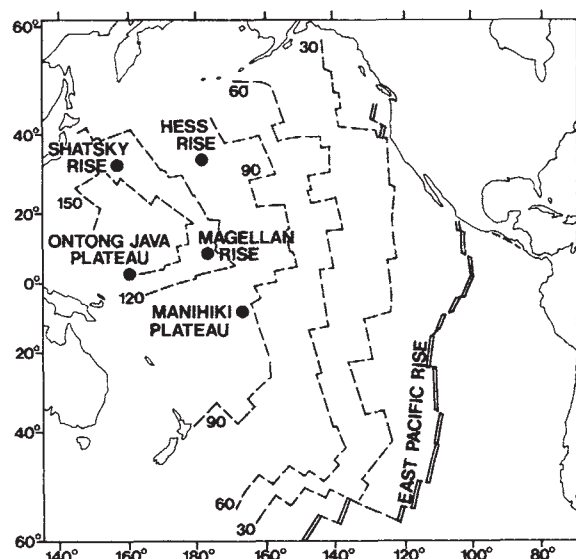


Fig. 1 Location of the larger oceanic plateaus in the Pacific with the age of the ocean floor in Myr.

fragments and evidence of shock metamorphism—are considerably more difficult to apply to the ocean bottom, particularly to a structure covered with abundant tholeiitic basalts and thick calcareous sediments. However, detailed geophysical surveys of plateaus may reveal remnants of arcuate structure and drilling in the central portion, dredging on the flanks and sampling in the surrounding sediments may reveal evidence for the petrographic effects of shock metamorphism. In addition, if large meteorite impacts prove to have widespread geochemical signatures²⁰, then the correlation of such signatures with the times of plateau formation would be very convincing evidence.

In light of the hypothesis put forward here, an obvious corollary question to ask is whether one of the large oceanic plateaus could be the topographic evidence of the Cretaceous–Tertiary boundary event. There is increasing evidence that this sudden and widespread extinction of the majority of species of flora and fauna on the Earth about 65 Myr ago was due to the

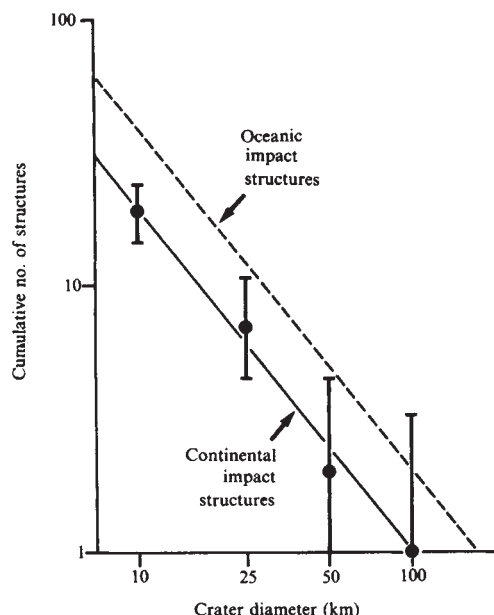


Fig. 2 The data points represent the cumulative number of probable impact structures on land greater than a given diameter for the last 150 Myr⁶. Error bars are 67% confidence limits for a poissonian distribution²⁴. The dashed line represents the number of structures expected in the ocean basins assuming the average age of the oceans is half the maximum of 150 Myr and that the area has been consistently three times that of the land mass.

impact of a giant meteorite^{11,12,20,21}. Preliminary investigation of ages in DSDP cores on the three largest oceanic plateaus, the Shatsky Rise, the Manihiki Plateau and the Ontong Java Plateau, rule them out because they are all older than 100 Myr^{3,7,9,16}. However, an oceanic impact marking this massive biological extinction could be represented by a large plateau that has now been subducted or coupled to a bordering continental margin. This plateau would have had to be much larger than the largest of the contemporary plateaus because there was not the same order of biological extinction at the time of initiation of formation of the three largest plateaus. The difficulty in subducting such a huge area of thickened crust should be a recognizable tectonic event in the ocean basins or in some of the ophiolite sequences at their margins. If the localized Cretaceous-Tertiary extinction of higher plants around the present day North Pacific²² surrounds the impact site¹¹, then the northwestern Pacific margin is a region in which to search for evidence of an anomalous tectonic event, occurring since the Cretaceous, that could represent the difficulty in subducting a massive plateau. A tectonic event of the appropriate magnitude might be the abrupt change in the motion of the Pacific plate about 45 Myr ago²³.

16. Douglas, R. G. & Moullade, M. *Bull. geol. Soc. Am.* **83**, 1163-1168 (1972).
17. Beaumont, C. *Geophys. J. R. astr. Soc.* **55**, 471-497 (1978).
18. Lambeck, K. & Nakiboglu, S. M. *J. geophys. Res.* **86**, 6961-6984 (1981).
19. Dence, M. R. 24th int. geol. Congr. Sect. 15, 77-89 (1972).
20. Alvarez, L. W., Alvarez, W., Asaro, F. & Michel, H. V. *Science* **208**, 1095-1108 (1980).
21. Smit, J. & Hertogen, J. *Nature* **285**, 198-200 (1980).
22. Hickey, L. J. in *Catastrophes in Earth History: The New Uniformitarianism* (Princeton University Press, 1982).
23. Morgan, W. J. *Mem. geol. Soc. Am.* **132**, 7-22 (1972).
24. Weichert, D. H. *Bull. seism. Soc. Am.* **70**, 1335-1346 (1980).

Volcanic ash deposits of early Eocene age from the Rockall Trough

E. J. W. Jones

Department of Geology, University College London, London WC1E 6BT, UK

A. T. S. Ramsay

Department of Geology, University College of Swansea, Singleton Park, Swansea SA2 8PP, UK

The stratigraphy of the Rockall Trough (Fig. 1) has proved difficult to elucidate because a mantle of Recent sediments usually prevents piston corers and dredges from sampling older accumulations. Although a seismic stratigraphy has been erected^{1,2}, considerable uncertainty surrounds the history of sedimentation and, consequently, the petroleum potential of the region. The upper 500-1,000 m of the succession are clearly Cenozoic deposits whose distribution has been strongly influenced by movements of Norwegian Sea overflow water¹. Deeper in the section, major current-controlled accumulations are absent, a feature observed elsewhere in the Atlantic¹. In view of the paucity of samples from these older deposits³, we report here the recovery of sediments that were laid down in the northern Rockall Trough before polar waters radically changed the depositional regime. The sediments record a period of explosive volcanicity during the early Eocene in the vicinity of the Wyville-Thomson Ridge.

The rock specimens were obtained from RRS *Shackleton* in 1979 during dredging operations at three locations in a trough separating the Wyville-Thomson Ridge and Faeroe Bank from Ymir Ridge (S1-S3, Fig. 1). This narrow depression acts as an important, although probably intermittent, entry point for dense Norwegian Sea overflow water into the Rockall Trough⁴. It was anticipated that the overflow had inhibited deposition locally and even caused sufficient erosion to expose formations normally deeply buried within the sediment pile. Figure 2 shows a complex pattern of current-controlled drift deposits associated with the overflow. These accumulations are absent south-west of kilometre 16 and along dredge tracks S1 and S2 on the side of Faeroe Bank, which is underlain by strongly reflecting, layered material pre-dating the drift sediments. A refraction profile on the crest of Faeroe Bank (V28-61; Fig. 1) suggests that the highly reflective sequence has a seismic velocity exceeding 4 km s⁻¹ and is therefore probably correlative with the 3.9-4.9 km s⁻¹ shallow volcanics of the Faeroes (Fig. 2).

Over 200 rocks, weighing a total of ~100 kg, were recovered at each dredge site (Fig. 1). The principal constituent is an indurated, manganese-stained, calcareous tuff, olive-green or grey in colour, containing a high proportion of unaltered glass shards. At sites S2 and S3 the tuffaceous components make up about 65% of the specimens. The striking lithological similarity between the tuffs indicates a local derivation from outcrops on the flanks of Faeroe Bank and Ymir Ridge. With the possible exception of some basaltic blocks, the remaining rocks at S2 and S3 consist of a diverse collection of glacial erratics. The dredge haul from S1 (Figs 1, 2) is more heavily contaminated with obvious glacial material.

Received 7 June; accepted 21 July 1982.

1. Den, N. *et al.* *J. geophys. Res.* **74**, 1421-1434 (1969).
2. Winterer, E. L., Lonsdale, P. F., Mathews, J. L. & Rosendahl, B. R. *Deep Sea Res.* **21**, 793-814 (1974).
3. Hussong, D. M., Wipperfurth, L. K. & Kroenke, L. W. *J. geophys. Res.* **84**, 6003-6010 (1979).
4. Carlson, R. L., Christensen, N. I. & Moore, R. P. *Earth planet. Sci. Lett.* **51**, 171-180 (1980).
5. Ben-Avraham, Z., Nur, A., Jones, D. & Cox, A. *Science* **213**, 47-54 (1981).
6. Grieve, R. A. F. & Robertson, P. B. *Icarus* **38**, 212-229 (1979).
7. Moberly, R. & Larsen, R. L. *Init. Rep. DSDP Leg 32*, 945-956 (1975).
8. Jackson, E. D. & Schlanger, S. O. *Init. Rep. DSDP No. 33*, 915-927 (1976).
9. Packham, G. H. & Andrews, J. E. *Init. Rep. DSDP No. 30*, 691-705 (1975).
10. Grieve, R. A. F. *Precamb. Res.* **10**, 217-247 (1980).
11. Emiliani, C., Kraus, E. B. & Shoemaker, E. M. *et al.* *Earth planet. Sci. Lett.* **55**, 317-334 (1981).
12. Napier, W. M. & Clube, S. V. M. *Nature* **282**, 455-459 (1979).
13. Dence, M. R., Grieve, R. A. F. & Robertson, P. B. in *Impact and Explosion Cratering*, 247-275 (Pergamon, New York, 1977).
14. Dence, M. R. & Grieve, R. A. F. *Lunar planet. Sci.* **10**, 282-294 (1979).
15. Grieve, R. A. F., Robertson, P. B. & Dence, M. R. *Proc. Lunar Planet. Sci.* **12A**, 37-57 (1981).

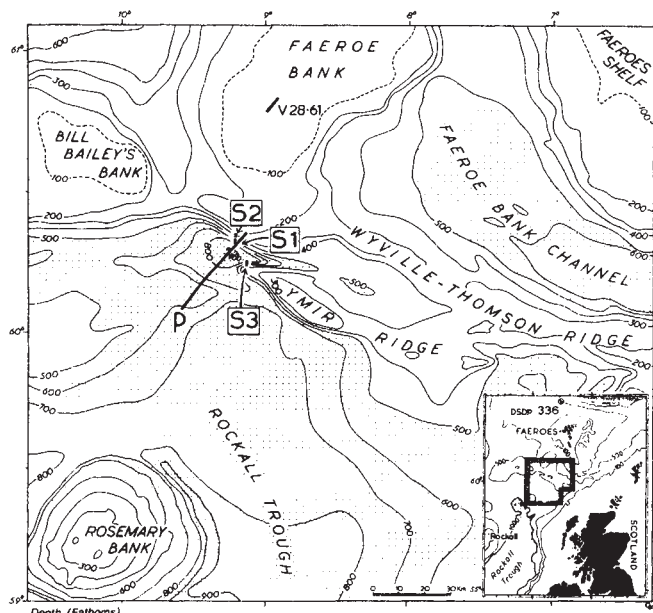


Fig. 1 The locations of Shackleton dredge sites S1–S3 at the northern end of the Rockall Trough. Bathymetric contours are based on a map of Ellett and Roberts⁷ and unpublished Shackleton soundings (100 fathoms = 183 m). The mid-points and depths of the sampling tracks are as follows; S1, 60°19.1' N, 9°13.2' W (986–1,450 m); S2, 60°20.3' N, 9°13.5' W (790–892 m); S3, 60°14.7' N, 9°08.6' W (1,223–1,602 m). An airgun reflection profile along track P is reproduced in Fig. 2. V28-61 is an unreversed seismic refraction profile shot with a sonobuoy receiver at the southern termination of the line. The refraction results are presented in Fig. 2.

The tuffaceous samples have been dated on the basis of calcareous nannofossils. All assemblages examined from the three dredge sites are Ypresian (early Eocene) in age and are assigned to the *Marthasterites tribrachiatus* zone of Brönnimann and Stradner⁵. The following species were recorded in each assemblage: *Braarudosphaera bigelowi*, *Chiasmolithus consuetus*, *Chiasmolithus grandis*, *Cyclococcolithus neogammation*, *Chiphragmalithus calathus*, *Discoaster barbadiensis*, *Discoasteroides kuepperi*, *Sphenolithus radians*, *Zygrhablithus bijugatus* and *Zygodolithus dubius*. The datum indicator *M. tribrachiatus* occurs in two samples from dredge S1, one specimen from S2 and five samples from S3. *Discolithina plana* and various species assigned to the genus *Micrantholithus* are found in most specimens. The presence of *B. bigelowi* and *Micrantholithus* indicates sedimentation at relatively shallow depths. The former is most abundant in the near-shore waters of the

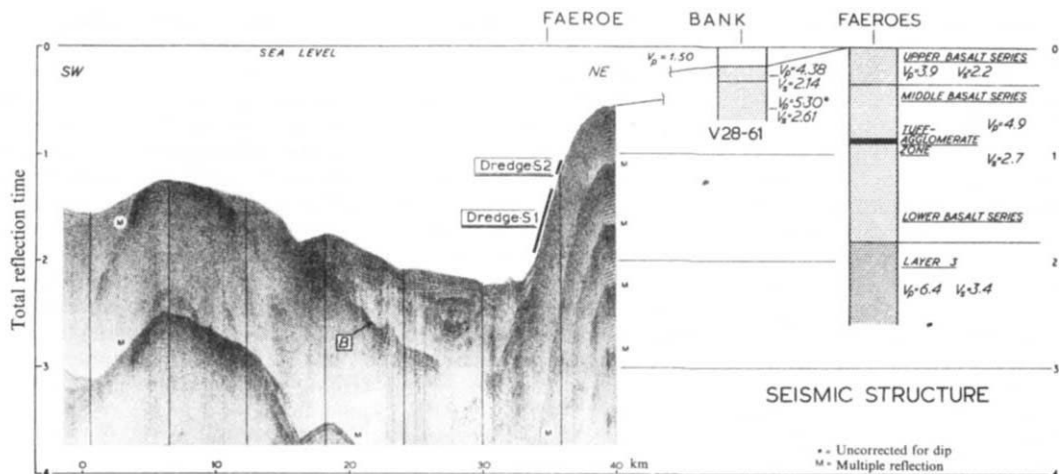
present Atlantic, while in Eocene sediments from the North Atlantic it is limited to deposits recovered from depths of less than 1,600 m (ref. 6).

The *M. tribrachiatus* zone is equivalent to NP12 of the standard calcareous nannoplankton zonation scheme proposed by Martini⁷. The numerical age of the zone is still a matter for debate. It is dated at 52–50 Myr BP by Hardenbol and Berggren⁸ (53–52 Myr adopting the 1976 ICC atomic abundance and decay constants). Odin and Curry⁹, however, argue that the zone is appreciably younger (49–47 Myr BP; ICC constants).

As the three dredges traversed a depth range of several hundred metres on both sides of the channel (Fig. 1), the restriction of the tuffaceous samples to zone NP12 suggests that a substantial thickness of Ypresian volcanogenic sediments outcrops in the area. Petrographic examinations and the major element contents of 19 specimens reveal compositions corresponding to tholeiitic basalts. The tuffs could have been derived from several volcanic regions which were active in the early Eocene. An oceanic rift to the west of the Faeroes and the Rockall Plateau is one possible area, as magnetic lineations of early Eocene age occur in the Greenland–Rockall Basin¹⁰. Rosemary Bank to the south¹¹, the Scottish Hebridean region and the Faeroese province, which we take to include the Wyville–Thomson Ridge, Faeroe Bank and Ymir Ridge, are other likely sources¹². A plot of the abundances of the elements Y, Zr and Ti, which are relatively insensitive to alteration processes¹³, shows the dredge samples clustering in the field normally occupied by within-plate basalts (Fig. 3). They are quite distinct from the basic lavas of Rosemary Bank¹¹ and from ocean-floor basalts, suggesting that Rosemary Bank and the oceanic rift to the west did not provide the volcanogenic material. The Hebridean area is an unlikely source of supply because winds with a strong easterly component probably prevailed in this region during the Lower Tertiary¹⁴. Local volcanic centres in the Faeroese province are therefore considered to be the provenance of the Ypresian tuffs reported here.

In the Faeroes a significant period of pyroclastic activity is recorded by the tuff-agglomerate zone separating the Lower from the Middle Basalt Series (Fig. 2)¹². According to Schilling and Noe-Nygaard¹⁵, the lavas below the tuff-agglomerate zone have compositions typical of plume basalts, whereas the volcanics above are of oceanic character, like the late Eocene basalts at Deep Sea Drilling Project (DSDP) Site 336 on the Faeroe–Iceland Ridge¹⁶ (Fig. 3). A palynological study by Laufeld¹⁷ suggests that the tuffs are Eocene in age, although they cannot be dated very precisely. As these are the only indications of substantial pyroclastic activity the tuff-agglomerate sequence may be coeval with the Ypresian volcanogenic sediments exposed on the flanks of Faeroe Bank and Ymir Ridge. The lower Eocene tuffs recovered by dredging may, as

Fig. 2 Shackleton dredge sites S1 and S2 shown in relation to the shallow seismic structure recorded using an airgun reflection profiler along track P in Fig. 1. Water-layer multiples are labelled M. The edge of Faeroe Bank is underlain by well stratified material which appears to outcrop on the southern slope. Between the 16 and 34 km positions on the horizontal scale, thick and irregular sedimentary drift deposits occur above the strong reflector B. The seismic structure of the upper portion of Faeroe Bank has been derived from refraction data recorded along track V28-61 in Fig. 1. The structure of the Faeroese volcanic pile is taken from the results of Pálmason²⁶. Seismic velocities are given in km s⁻¹.



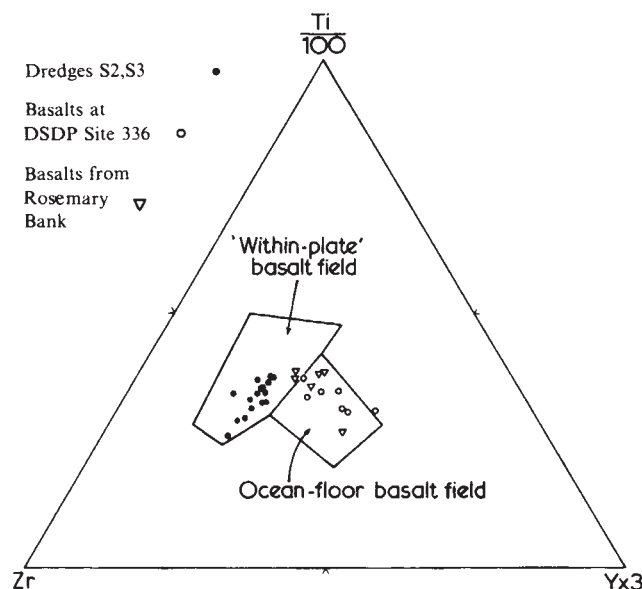


Fig. 3 Trace element variations in the Ypresian tuffs from dredge sites S2 and S3 shown on a Ti-Zr-Y discrimination diagram. Compositional fields are derived from Pearce and Cann¹³. The tuffs are quite distinct from the basaltic rocks dredged from Rosemary Bank¹¹ and from ocean-floor basalts such as the late Eocene lavas drilled at DSDP Site 336¹⁶ north of the Faeroes (Fig. 1).

in the Faeroes, be interbedded with thick basalt units, for seismic reflection profiles show layering beneath Faeroe Bank and Ymir Ridge, while high-amplitude, short-wavelength magnetic anomalies^{18,19} imply strongly magnetic, probably basaltic rocks, close to the seabed.

Correlation can be made further afield with the Rockall Plateau and the North Sea where Lower Tertiary ash falls are recorded. An early Tertiary ash sequence, much of it derived from basaltic sources, is widespread in the North Sea and forms an important seismic marker²⁰. The same tuffaceous interval also appears in wells recently drilled west of the Shetlands²¹. According to Jacqu  and Thouvenin²⁰, the climax of the pyroclastic activity occurred in the late Palaeocene, at the time of deposition of the Woolwich Formation in southern England (zone NP9). The dredged tuffs are somewhat younger, but Ypresian ash layers have been encountered in wells in the northern North Sea²⁰. In East Anglia, over 40 thin ash beds of Ypresian age, considered by Jacqu  and Thouvenin to be slightly younger than the main North Sea tuffs, have been discovered at the base of the London Clay^{22,23}. These are believed to be contemporaneous with ash bands in the Mo Clay Formation of Denmark²⁴. Furthermore, Harrison *et al.*²⁵ have found tuffaceous sediments in the late Palaeocene-early Eocene (NP9 to NP12) sections drilled on the western side of the Rockall Plateau, thus extending considerably the known limits of the Lower Tertiary ash falls. The volcanogenic deposits from Faeroe Bank and Ymir Ridge clearly form part of the widespread ash series which accumulated around Britain during late Palaeocene-early Eocene time. The provenance of these extensive ash accumulations is problematical. However, as Faeroe Bank and Ymir Ridge appear to be underlain by thick pyroclastic sequences of Ypresian age, the Faeroese province was probably an important source for an appreciable proportion of the lower Eocene volcanic ash in the North Sea and surrounding areas.

We thank the officers, crew and scientific party aboard RRS *Shackleton*, the staff of the NERC Research Vessel Services, Robert Houtz of the Lamont-Doherty Geological Observatory and Professor Dennis Curry for their help in this study, and Anthony Osborn for carrying out chemical analyses. Financial support was provided by the NERC.

Received 19 January; accepted 19 July 1982.

1. Jones, E. J. W., Ewing, M., Ewing, J. I. & Eittreim, S. L. *J. geophys. Res.* **75**, 1655-1680 (1970).
2. Roberts, D. G. *Phil. Trans. R. Soc. A* **278**, 447-509 (1975).
3. Jones, E. J. W., Ramsay, A. T. S., Preston, N. J. & Smith, A. C. S. *Nature* **251**, 129-131 (1974).
4. Ellett, D. J. & Roberts, D. G. *Deep-Sea Res.* **20**, 819-835 (1973).
5. Br nnimann, P. & Stradner, H. *Erd l-Z. Bohr- u. F rdertech.* **76**, 364-369 (1960).
6. Ramsay, A. T. S. *Nature* **236**, 67-70 (1972).
7. Martini, E. *Proc. 2nd int. Conf. Planktonic Microfossils*, Rome 739-805 (Tecnoscienza 1971).
8. Hardenbol, J. & Berggren, W. A. *Contributions to the Geologic Time Scale* (eds Cohee, G. V., Glaessner, M. F. & Hedberg, H. D.) 213-234 (American Association of Petroleum Geologists, Tulsa, 1978).
9. Odin, G. S. & Curry, D. C. *r. hebd. S anc. Acad. Sci., Paris* **293**, 1003-1006 (1981).
10. Vogt, P. R. & Avery, O. E. *J. geophys. Res.* **79**, 363-389 (1974).
11. Dietrich, V. J. & Jones, E. J. W. *Mar. Geol.* **35**, 287-297 (1980).
12. Rasmussen, J. & Noe-Nygaard, A. *Geology of the Faeroe Islands* (Geological Survey of Denmark I, Ser. No. 25, Copenhagen, 1970).
13. Pearce, J. A. & Cann, J. R. *Earth planet. Sci. Lett.* **19**, 290-300 (1973).
14. Donn, W. L. & Ninkovich, D. J. *geophys. Res.* **85**, 5455-5460 (1980).
15. Schilling, J.-G. & Noe-Nygaard, A. *Earth planet. Sci. Lett.* **24**, 1-14 (1974).
16. Talwani, M. & Shipboard Scientific Party Init. Rep. *DSDP Leg 38*, 23-116 (US Govt Printing Office, Washington DC, 1976).
17. Laufeld, S. *Geol. F oreningens Stockholm F rhand.* **87**, 231-238 (1965).
18. Bott, M. H. P. & Stacey, A. P. *Deep-Sea Res.* **14**, 7-11 (1967).
19. Avery, O. E., Burton, G. D. & Heirtzler, J. R. *J. geophys. Res.* **73**, 4583-4600 (1968).
20. Jacqu , M. & Thouvenin, J. in *Petroleum and the Continental Shelf of North-West Europe*, I (ed. Woodland, A. W.) 455-465 (Applied Science Publishers, London, 1975).
21. Ridd, M. F. in *Petroleum Geology of the Continental Shelf of North-West Europe*, (eds Illing, L. V. & Hobson, G. D.) 414-425 (Institute of Petroleum, London, 1981).
22. Elliott, G. F. *Nature phys. Sci.* **230**, 9 (1971).
23. Knox, R. W. O'B. & Harland, R. J. *geol. Soc. Lond.* **136**, 463-470 (1979).
24. Pedersen, A. K., Engell, J. & R nsbo, J. G. *Lithos* **8**, 255-268 (1975).
25. Harrison, R. K., Knox, R. W. O'B. & Morton, A. C. *Init. Rep. DSDP Leg 48*, 771-785 (1979).
26. P lmason, G. *Tectonophysics* **2**, 475-482 (1965).

Evidence for earlier date of 'Ubeidiya, Israel, hominid site

Charles A. Repenning

United States Geological Survey, Menlo Park, California 94025, USA

Oldrich Fejfar

Geological Survey of Czechoslovakia, Praha, Czechoslovakia

It has been suggested that the evolution of man took place in Africa. This suggestion results from the unusual abundance of fossil material in Africa that is quite ancient in comparison with what is known elsewhere. The theory of an African origin has influenced the interpretation of the age of some non-African archaeological sites. A case in point is the 'Ubeidiya locality in Israel, which is generally considered to be about 700,000 yr old because it has been assumed by a few that the associated Early Acheulian tool industry, and the persons who used it, would have taken considerable time to disperse from Olduvai Gorge to this non-African site in Israel. Here we evaluate fossil mammals from 'Ubeidiya, which are stratigraphically and directly associated with Early Acheulian artefacts, and find no substantial reason for considering the locality younger than 2 Myr, and possibly as much as 500,000 yr older than any record of Early Acheulian artefacts or *Homo erectus* in Africa.

Abundant Early Acheulian artefacts¹ with associated extinct mammals have been known from the 'Ubeidiya area, 3 km south of the Sea of Galilee on the west side of the valley of the Jordan River, since agricultural clearing exposed the site in 1959^{2,3}. A few fragmentary remains have been referred to *Homo erectus*, although Tobias⁴ felt that they were too incomplete for specific designation. Haas⁵ has described the mammalian fauna from 'Ubeidiya, which is most recently summarized by Tchernov⁶. The site is usually considered to be older than 700,000 yr but probably not much older^{7,8}.

Early Acheulian artefacts and remains of *Homo erectus* first occur in Olduvai Gorge, Tanzania, in the deposits above the

Lemuta Member in the lower part of Bed II. The Lemuta Member is dated as about 1.5 Myr in age. In Koobi Fora, Kenya, ~800 km closer to Israel, skull KNM-ER 3733 appears to predate the 1.5 Myr Okote Tuff⁹. The abrupt appearance of *Homo erectus* and Early Acheulian artefacts in Bed II of Olduvai has suggested to some the possibility of an invasion¹⁰. "Most, if not all, students consider it likely that *H. erectus* gave rise to *H. sapiens*."¹¹

Although many authors accept the African origin of man the toolmaker, others have suggested that the development of man took place in more temperate environments¹². The question of how much older than 700,000 yr the 'Ubeidiya Acheulian archaeological record thus becomes of interest: is it older than Bed II in Olduvai Gorge and, if so, is it possible that *Homo erectus* evolved elsewhere than Africa and invaded Africa more than 1.5 Myr ago^{12?}

As listed by Tchernov⁶, the 'Ubeidiya fauna contains several mammals that present a 'late Villafranchian' aspect, comparable with that at Seneze, France¹³, as well as that at St Vallier, France, and Kislang, Hungary. In the micromammal biochronology of Europe¹⁴, it is a fauna characteristic of the Villanyian age which has been dated as being at least as old as 2.5 Myr (Roca Neyra, France) and at least as young as about 1.9 Myr (Le Coupet, France). (There is and will probably always be inconsistency in the use of the term Villafranchian age. In micromammal chronology the latest Villafranchian is separable as a distinct mammalian age, the Villanyian.) The Seneze fauna itself is from the top of a sequence of lake sediments overlying a basalt dated 2.3 Myr old and is immediately upsection from one of the Reunion polarity events¹⁵, so the fauna must be between 2.12 and 2.0 Myr old. A Villanyian age assignment corresponds approximately to some previous mammalian age assignments for the 'Ubeidiya fauna^{5,16}, but reflects subsequent refinement and temporal calibration of the biochronology (see Fig. 1). Extraordinary support for the calibration of the temporal limits of the Villanyian micromammal age (as used by Fejfar and Heinrich¹⁴) comes from the correlation of Holarctic microtine rodent dispersal events that mark the beginning and end of this age¹⁷.

The correlation of the 'Ubeidiya fauna with the Villanyian age of Europe is very good, although not perfect. On the basis of Tchernov's listing⁶, the following mammals of the 'Ubeidiya fauna are of particular significance.

The zebnine horse, *Equus stenonis*, is reported from 'Ubeidiya as the earliest record from Israel; it is first known in Europe from 'late Villafranchian' (or Villanyian) faunas and the earliest occurrence is in the Roca Neyra fauna (2.5 Myr). The earliest East African records of zebnine horses are 2.01–2.14 Myr old^{18,19}, and much older than the first record of Early Acheulian artefacts and *Homo erectus*, even though they indicate a later immigration than that of Europe. The zebras had to pass 'Ubeidiya en route to Africa and dispersed rapidly across the continent¹⁸.

The pig *Sus* sp. cf. *Sus strozzii* is a typical 'late Villafranchian' or Villanyian form and is believed to derive from *Sus minor* of the early Pliocene of Europe (earlier Villafranchian or older). *S. strozzii* was replaced by *Sus scrofa* (the modern domestic pig) in the early Biharian faunas of Europe. In Israel, *S. strozzii*, but no zebnine horses, is also present in the older Bethlehem fauna⁶ and its presence in 'Ubeidiya could have been as a relict, but *S. scrofa* also replaced *S. strozzii* in younger faunas of Israel⁶ and it dispersed westward along the southern shore of the Mediterranean where it is present in the Ternifine fauna, Algeria²⁰.

Cervid artiodactyls (deer and elk) appear in Israel for the first time in the 'Ubeidiya fauna^{5,6,21}. They are present in this fauna in a greater number of specimens than any other large mammal except *Hippopotamus*⁵ and are assigned to four forms. *Cervus* sp. cf. *Cervus ramosus* (= *Croizetoceros ramosus* or *Anoglochis ramosus*) is a species first known from the Villaroya fauna of Spain (Villanyian age) and possibly last known from the Saint Esteve–Janson fauna of France (less than 450,000 yr

old. *Euctenoceros senegensis* was described from the Seneze fauna and is also known from the Roca Neyra fauna of France (2.5 Myr old); it is possibly last known from the Cromer Forest Bed, England (600,000–730,000 yr old). *Dama* sp. is known from the Villanyian faunas of Europe into the recent fauna. The oldest (Villanyian? equivalent) *Dama* known from Africa occurs with zebnine horses near Wadi Halfa, Sudan²². *Megaloceros* sp. (giant deer) is reported on the basis of a single large deciduous last premolar⁵ of questionable value in identification. The genus is first known without question from European early Biharian faunas but is questionably present in faunas as old as Kislang.

Bison priscus (steppe wisent), to which two tooth fragments are assigned⁵, is known in Europe only from early Biharian and younger faunas but is reported from Villanyian-equivalent faunas in Siberia²³. Its distinction from the Villafranchian genus *Leptobos* on two tooth fragments is questioned.

Crocota crocuta (spotted hyaena) is first known in latest Biharian faunas of Europe. However, *Crocota* sp., often referred to as *Crocota crocuta*, is present in African deposits at least as old as Member G of the Shungura Formation²⁴, Ethiopia, and Bed I at Olduvai, Tanzania. These records are as old as those of zebnine *Equus* in Africa.

Ursus etruscus (primitive bear) is primarily, or entirely, known from Villanyian faunas including Seneze, France. It derives from an earlier Villafranchian form and is ancestral to Biharian bears of Europe and Asia, one of which, *Ursus arctos*, is known from younger faunas in Israel and westward along the northern coast of Africa to the Ternifine fauna, Algeria, and in Morocco. The specific identity of the 'Ubeidiya material is dependent on Tchernov's⁶ note that the premolars have a primitive construction.

Megantereon megantereon (sabre- or dirk-tooth tiger) is a form characteristic of the Villafranchian in Europe and the youngest records are of Villanyian age, although the lineage persisted later in Asia and survived in North America until the early Holocene. *Megantereon* is known from late Pliocene (Villanyian) faunas throughout Africa and is reported in East Africa²⁴ in members B to G of the Shungura Formation of Ethiopia.

The micromammals of the 'Ubeidiya fauna also indicate a Villanyian age; 18 (ref. 5) to 20 (ref. 6) forms have been identified in the fauna. In adjacent articles Janossy¹⁶ and Jaeger²⁵ provide interpretations of the age of the 'Ubeidiya micromammals. Jaeger indicates that the absence of the Quetta mole-vole, *Ellobius fuscocapillus*, and the grass rat, *Arvicanthis ectos* (an African immigrant), in the 'Ubeidiya fauna and their presence in the younger Oumm Quatafa fauna of Israel and in the Ternifine fauna of Algeria indicate that the 'Ubeidiya fauna is older than that from Ternifine, a suggestion noted above in connection with the presence of *S. scrofa* and *U. arctos* from Ternifine.

Janossy¹⁶ presented an outline of the micromammal biochronology of Europe, on the basis of microtine rodents and soricid insectivores. He pointed out that the 'Ubeidiya fauna of Israel correlated with the beginning of the Biharian micromammal age (as then used), and with the Brielle fauna, Netherlands (which correlates with the earliest Biharian as now used, as well as in 1975 usage).

Of the many micromammals in the 'Ubeidiya fauna, only *Hypolagus* sp. (primitive hare), *Jordanomys pusillus* and *Lagurodon arankae* (primitive steppe lemmings) need discussion here. Around the Holarctic world the primitive archaeolagine *Hypolagus* is replaced by advanced leporids, primarily *Lepus* (modern hare), when these appear. In North America large species of *Hypolagus* vanish, apparently instantaneously, with the introduction of advanced leporids from Asia ~2.5 Myr ago. In Europe *Hypolagus* survives at least until the early part of the Biharian micromammal age in such faunas as Kamyk, Poland, that are more or less the temporal equivalent of the Brielle fauna of the Netherlands and the Olduvai event (~1.7–1.9 Myr ago). In Africa, *Hypolagus* is not known, but

Howell for his remarkable efforts in providing a technical review of a preliminary draft. L. L. Jacobs, K. R. Lajoie, E. S. Lowell, J. D. Lowell and H. Merrick also reviewed the manuscript and

Received 6 May; accepted 21 July 1982.

1. Bar-Yosef, O. in *After the Australopithecines* (eds Butzer, K. W. & Isaac, G. L.) 571-604 (Mouton, The Hague, 1975).
2. Stekelis, M. *Israel Acad. Sci. Humanities* (1966).
3. Stekelis, M., Bar-Yosef, O. & Schick, T. *Israel Acad. Sci. Humanities* (1969).
4. Tobias, P. V. *Nature* **211**, 130-133 (1966).
5. Haas, G. *Israel Acad. Sci. Humanities* (1966).
6. Tchernov, E. in *The Quaternary of Israel* (ed. Horowitz, H.) Sect. 7 (Academic, New York, 1979).
7. Siedner, G. & Horowitz, A. *Nature* **250**, 23-26 (1974).
8. Horowitz, H. (ed.) *The Quaternary of Israel* (Academic, New York, 1979).
9. Walker, A. & Leakey, R. E. F. *Scient. Am.* **239**(2), 54-66 (1978).
10. Leakey, M. D. *Olduvai Gorge 3. Excavations in Beds I and II, 1960-1963*, 272 (Cambridge University Press, 1971).
11. Howell, F. C. in *Evolution of African Mammals* (eds Maglio, V. J. & Cooke, H. B. S.) 154-248 (Harvard University Press, Cambridge, 1978).
12. Kretzoi, M. *Anthrop. kozl.* **20**, 3-11 (1976).
13. Schaub, S. *Eclog. geol. Helv.* **36**, 270-289 (1943).
14. Fejfar, O. & Heinrich, W.-D. *Eclog. geol. Helv.* **74**, 997-1006 (1981).
15. Prevot, M. & Dalrymple, G. B. C. *Arch. Séanc. Acad. Sci., Paris* **271**, 2221-2224 (1970).

provided many helpful suggestions. Marion E. Anderson typed the manuscript.

16. Janossy, D. in *After the Australopithecines* (eds Butzer, K. W. & Isaac, G. L.) 375-397 (Mouton, The Hague, 1975).
17. Reppening, C. A. *Can. J. Anthropol.* **1**, 37-44 (1980); *Acta zool. fennica* (in the press).
18. Churcher, C. S. & Hooijer, D. A. *Zoöl. Meded., Leiden* **55**, 265-280 (1980).
19. Mankinen, E. A. & Dalrymple, G. B. *J. geophys. Res.* **84**, 615-626 (1979).
20. Cooke, H. B. S. & Wilkinson, A. F. in *Evolution of African Mammals* (eds Maglio, V. J. & Cooke, H. B. S.) 435-482 (Harvard University Press, Cambridge, 1978).
21. Bar-Yosef, O. & Tchernov, E. *Israel Acad. Sci. Humanities* (1972).
22. Lydekker, R. Q. *Jl geol. Soc. Lond.* **43**, 161-164 (1887).
23. Vangengeim, E. A. *Acad. Sci. U.S.S.R. Geol. Inst.* (1977).
24. Howell, F. C. in *Early Hominids of Africa* (ed. Jolly, C. J.) 85-130 (Duckworth, London, 1978).
25. Jaeger, J.-J. in *After the Australopithecines* (eds Butzer, K. W. & Isaac, G. L.) 399-418 (Mouton, The Hague, 1975).
26. Cooke, H. B. S. in *African Ecology and Human Evolution* (eds Howell, F. C. & Bouliere, F.) 65-116 (Aldine, Chicago, 1963).
27. Corbet, G. B. *The Mammals of the Palaearctic Region* (British Museum (N.H.), London, 1978).
28. Topachevskiy, V. O. *Acad. Sci. Ukrain. S.S.R., Inst. Zool.*, 1-164 (1965).
29. Tchernov, E. *Věd. naturf. Ges. Basel* **79**, 161-185 (1968).
30. Kretzoi, M. *Föld. Intézet, Evi jelentései*, 239-265 (1954).

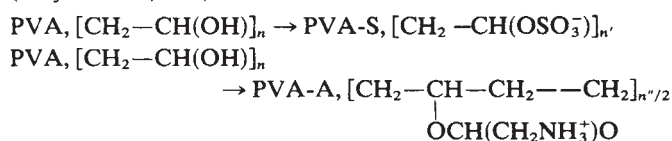
A protective function of the coacervates against UV light on the primitive Earth

Hiroyuki Okihana & Cyril Ponnampерума*

Laboratory of Chemical Evolution, University of Maryland,
College Park, Maryland 20742, USA

Results from simulation experiments in the laboratory support the Oparin-Haldane hypothesis of chemical evolution¹⁻⁴. The organic compounds produced may have accumulated in the primitive sea, constituting the so-called 'primordial soup' of organic molecules. Beyond this primordial stage, however, many developments were required before the emergence of the first cellular organism. The first cell need not necessarily have been a modern cell, complete with a membrane, a chromosome, ribosomes, enzymes, a metabolism, and the property of self-replication; the primary requirement of the protocell would have been the ability to evolve into a complete cell. There are many models of the protocell: for example coacervate droplets⁵⁻⁹, proteinoid microspheres^{10,11}, micelles (lipid vesicles)¹², and marigranules^{13,14}. Of these, the coacervate droplets are of interest as they are dynamic^{6,7} and because their mechanism of formation is perhaps better understood¹⁵⁻¹⁶. Coacervate droplets are essentially small bounded spaces in which the concentration of substances is higher than in the surrounding medium and in which the distribution of the substances differs from that in the solutions. Their essential functions in the protocell model are: (1) selection; (2) concentration; (3) production of hypohydrous effects; (4) interaction; and (5) organization⁵⁻⁹. Here we report another possible function of the coacervate. We added glycine or diglycine to the coacervate system, then irradiated the system with an argon lamp. We report that the coacervate droplets concentrated glycine or diglycine, thus protecting them from decomposition by irradiation. These results suggest a protective function of the coacervate against UV light on the primitive Earth.

Partially sulphated polyvinyl alcohol (PVA-S) and a partially amino-acetylsed alcohol (PVA-A) were obtained by the esterification¹⁷ of commercially available polyvinyl alcohol (Polysciences, Inc.):



The products were dialysed against water and deionized by

passage through mixed-bed Rexyn I-300 ion-exchange resin (Fisher Science). The degree of esterification was estimated by conductimetric titration¹⁷. One of the six PVA-A and one of the five PVA-S compounds were used in this study. The molecular weights of the PVAs from which the PVA-A-6 and PVA-S-1 compounds were derived were 3,000 and 78,000, respectively. The degrees of esterification were 4.99 mol % and 34.1 mol %, respectively.

PVA-A-6 and PVA-S-1 were mixed at equal charge density ratio in sodium phosphate buffer (0.025 M, pH 7.5) at room temperature. On mixing of the two compounds, the solution became turbid and coacervate droplets were observed under the microscope. Some of them coagulated into larger ones having vacuoles and others assembled to form colonies.

Glycine or diglycine (0.2 μ mol) was added to 1 ml of the system at the time of formation of the coacervate droplets (experiment 1) or after their formation (experiment 2). After incubation for 30 min, the coacervate droplets were separated from the surrounding medium (equilibrium liquid) by low-speed centrifugation (3,000 r.p.m. for 10 min). In experiment 2, the mixture of PVA-A-6 and PVA-S-1 was incubated in buffer for 10 min before the addition of glycine or diglycine. Glycine and diglycine were concentrated more in the coacervate droplets than in the surrounding equilibrium liquid; that is, about three times more for glycine and about twice for diglycine (Table 1). There were no differences between the two experiments. Equilibrium between the coacervate droplets and the equilibrium liquid was completed within the 30 min of incubation.

Concentration of substances from the surrounding medium is one of the most essential functions of the coacervate droplets. The droplets used in this study were the simplest ones, that is, having no membranes at the surface nor any enzymes within. The concentrations of glycine and diglycine achieved must, therefore, be due to their solubilities between the concentrate phase (the coacervate droplets) and the dilute phase (the equilibrium liquid), that is, the partition equilibrium, and not due to any active transport on the part of the coacervate droplets.

As an irradiation source for the photochemical decomposition experiments, a high pressure argon lamp¹⁸ (15 atm; Gianini, Photon Flux) (UV range from $\sim 5 \times 10^{16}$ to 10^{18} with visible range at $\sim 5 \times 10^{18}$ photons $\text{cm}^{-2}\text{s}^{-1}$, 50 Å) was used because of the similarity of its spectrum to that of the Sun. The coacervate droplets were separated by low-speed centrifugation, washed once with buffer without both PVA-A-6 and PVA-S-1, and were used in the following experiments. The reaction mixture consisted of 0.025 M sodium phosphate buffer pH 7.5 and 0.5 mM of glycine or diglycine, with or without the coacervate droplets. Quartz cells for UV spectroscopy (1 cm $\times$ 1 cm) were used as the irradiation cells. These were attached to an emission window of the argon lamp. The systems were closed with caps, and were gently agitated by convection during the irradiation.

*To whom correspondence should be addressed.

The effects of irradiation on glycine and diglycine are shown in Fig. 1. For the first 2 h, the reaction mixtures were held in the reaction cells without any irradiation; no decomposition occurred. Then the samples containing glycine and diglycine were irradiated for 4 h and 1 h, respectively. In the systems without the coacervate droplets, 84% of the glycine decomposed after 4 h of irradiation (Fig. 1b), while 97% of the diglycine decomposed in 1 h (Fig. 1f). In the coacervate droplets

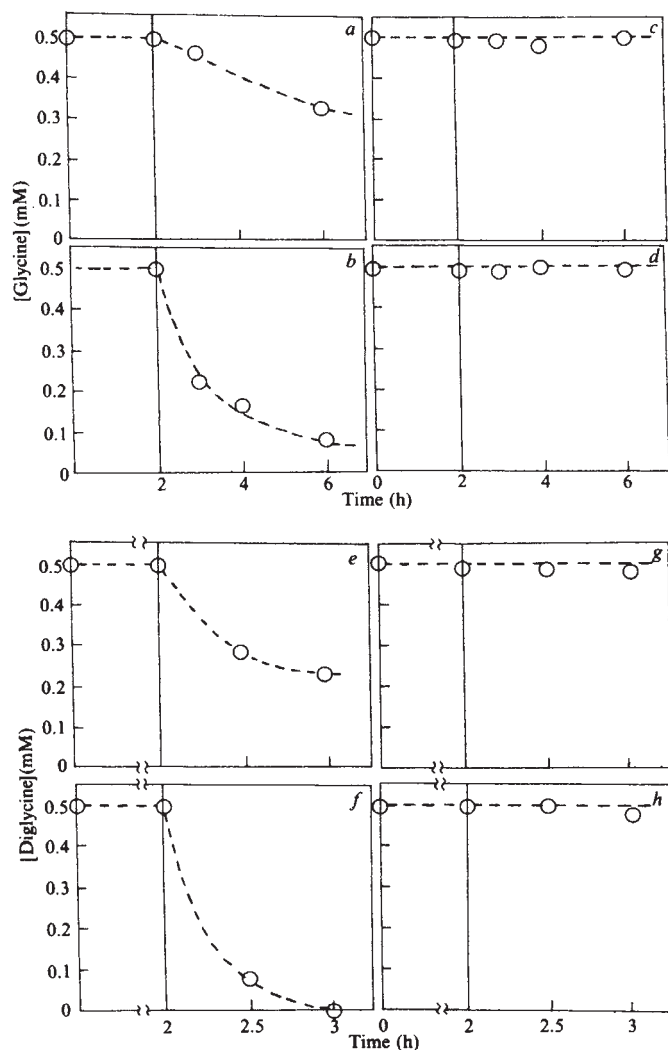


Fig. 1 *a-d*, Time course of glycine irradiation. The concentrations of glycine in the systems were plotted against time. Every sample was kept in the absence of irradiation for the first 2 h. The samples with (*a*) and without (*b*) coacervate droplets were irradiated for 4 h. The same samples with (*c*) and without (*d*) coacervate droplets were also kept in the absence of irradiation. *e-h*, Time course of irradiation of diglycine. Except for the irradiation time, conditions were the same as described for glycine.

systems, 34% of glycine (Fig. 1*a*) and 52% of diglycine (Fig. 1*e*) decomposed within the same periods of irradiation.

The decomposition processes appeared to be those of a first-order reaction (see Fig. 2). The kinetic constants were estimated from the slopes of $\ln(A_0/A_t)$ plotted against time, where A_t and A_0 represent the amounts of reactants at time t and at $t = 0$, respectively. The constants for reactions without the coacervate droplets were 0.44 h^{-1} for glycine and 3.5 h^{-1} for diglycine. With the coacervate droplets both constants decreased to 0.093 h^{-1} and 1.8 h^{-1} for glycine and diglycine, respectively.

The products of the diglycine solution were analysed after 1 h of irradiation. The analyser was equipped with a normal ninhydrin detection system so that only compounds containing an NH_2 group were detected. As judged from the retention time, 0.24 mM of diglycine, 0.16 mM of methylamine, 0.071 mM of ammonia and 0.029 mM of α -aminobutyric acid were detected in the case of the coacervate droplets system, and 0.015 mM of diglycine, 0.11 mM of methylamine, 0.061 mM of ammonia and 0.026 mM of α -aminobutyric acid were detected in the absence of the coacervate droplets system. There were no peaks at the position of glycine. Some other small peaks were present which have not been identified. The photochemical products of diglycine are quite different from the thermal hydrolysed ones.

Although the detailed photochemical mechanisms are unknown, the coacervate droplets protected glycine and digly-

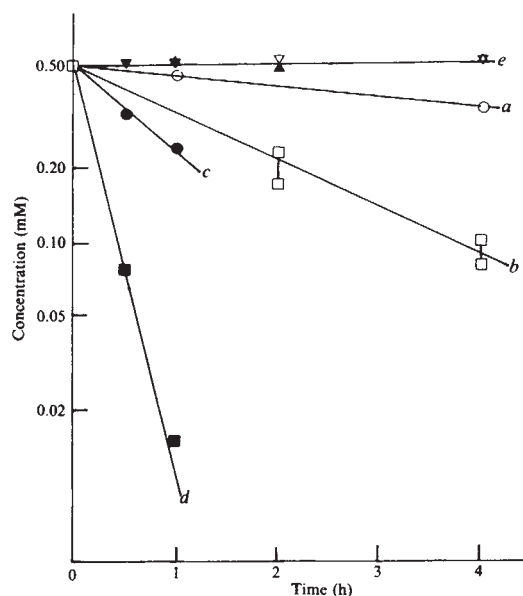


Fig. 2 Semi-log plots of the reactions. *a*, Glycine, with coacervate droplets; *b*, glycine without coacervate droplets; *c*, diglycine with coacervate droplets; and *d*, diglycine without coacervate droplets (shown in Fig. 1*a*, *b*, *e*, and *f*, respectively). Line *e* represents all the experiments without irradiation. Δ , Gly; $\blacktriangle$, gly with coacervate; $\star$, digly; $\blackstar$, digly with coacervate.

Table 1 Concentrations of glycine and diglycine in the experimental systems

		Equilibrium liquid			Coacervate droplets			
		Concentration (mM)	Volume (ml)	amount (μmol)	Concentration (mM)	Volume (ml)	amount (μmol)	Concentration ratio
Expt 1	Gly	0.164 ± 0.02	0.87	0.143	0.435 ± 0.01	0.13	0.057	2.7
	(Gly) ₂	0.169 ± 0.02	0.82	0.139	0.340 ± 0.1	0.18	0.061	2.0
Expt 2	Gly	0.167 ± 0.02	0.89	0.149	0.479 ± 0.06	0.11	0.053	2.9
	(Gly) ₂	0.171 ± 0.01	0.84	0.144	0.356 ± 0.01	0.16	0.057	2.1

Glycine or diglycine ($0.2 \mu\text{mol}$) were added to the systems (1) at the time of coacervate formation or (2) after their formation. Total volume of each of the systems was 1 ml. Amounts of glycine or diglycine are the product of the concentration and the volume.

cine from decomposition by irradiation. There are two possible explanations for this effect: (1) the coacervate droplets concentrated glycine or diglycine from the surrounding medium. These concentrated substances would be much less susceptible to irradiation, decomposing considerably less than those in the homogeneous solution or in the surrounding medium. (2) The coacervate droplets screened light, because transmitted light through the coacervate system was much less intensive than that through the homogeneous solution. Glycine and diglycine even in the surrounding medium were decomposed less than in the homogeneous solution. Thus, equilibrium is established between the coacervate droplets and the surrounding medium. As the decomposition of glycine or diglycine in the surrounding medium occurs, these substances are released from the coacervate droplets into the surrounding medium and vice versa.

We do not propose that the very same molecules as the

synthetic polyelectrolytes used here might have existed in the primitive sea, but they represent compounds having similar molecular weight and charge density. The molecular weights and charge densities of PVA-S-6 and PVA-S-1 are similar to those of some proteins, and of DNA or RNA. It is reasonable to assume that polyelectrolytes in the primitive sea, such as oligonucleotide-like polymers and oligopeptide-like polymers could have interacted much like the synthetic ones used in this study.

Water in the primitive sea has been said to protect the ingredients in the primordial soup from decomposition by irradiation. The coacervate droplets enhanced this effect and also selectively concentrated these ingredients. This function might have been one of the most important attributes of primitive microsystems, leading to further development and ultimately to the first living organism.

Received 20 April; accepted 16 June 1982.

1. Miller, S. L. *Science* **117**, 528–529 (1953).
2. Ponnampetuma, C. *Cosmochemistry* (ed. Milligan, W. O.) 137–197 (Robert A. Welch Foundation, Houston, 1978).
3. Miller, S. L. & Orgel, L. E. *The Origins of Life on the Earth* (Prentice-Hall, New Jersey, 1974).
4. Dillan, L. S. *The Genetic Mechanism and the Origin of Life* Ch. 1 (Plenum, New York, 1978).
5. Bungenberg de Jong, H. G. *Colloid Science* Ch. 10 (Elsevier, Amsterdam, 1949).
6. Oparin, A. I. *The Origin of Life* (Macmillan, New York, 1938).
7. Evreinova, T. N. *Coacervates* (Kyoritsu, Tokyo, 1974).
8. Mizutani, H., Okihana, H., Hasegawa, M., Yano, T. & Noda, H. *The Origin of Life and Evolutionary Biochemistry* (eds Dose, K., Fox, S. W., Deborin, G. A. & Pavlovskaya, T. E.) 339–354 (Plenum, New York, 1974).
9. Noda, H., Mizutani, H. & Okihana, H. *Origins of Life* **6**, 139–146 (1975).
10. Fox, S. W., Harada, K. & Kendrick, J. *Science* **129**, 1221–1223 (1959).
11. Fox, S. W. *The Origins of Prebiological Systems and of Their Molecular Matrices* (ed. Fox, S. W.) 361–373 (Academic, New York, 1965).
12. Goldacre, R. J. *Surface Phenomena in Chemistry and Biology* (eds Danielli, J. R., Pankhurst, K. G. A. & Riddiford, A. C.) 276–278 (Pergamon, London, 1958).
13. Yanagawa, H. & Egami, F. *Proc. Japan Acad.* **53**, 42–45 (1977).
14. Okihana, H. *Viva Origino* **8**, 19–22 (1979).
15. Okihana, H. & Noda, H. *Origin of Life* (ed. Noda, H.) 375–383 (Japan Scientific Press, Tokyo, 1978).
16. Tainaka, K. *Biopolymers* **19**, 1289–1298 (1980).
17. Okihana, H. & Nakajima, A. *Bull. Inst. chem. Res.* **54**, 63–71 (1976).
18. Rushbrook, P., Ponnampetuma, C. & Reinisch, R. F. *Abstr. Pacific Conf. Chemistry and Spectroscopy*, Am. Chem. Soc. Meet., San Francisco (1970).

Phenotypic evolution in a poorly dispersing snail after arrival of a predator

Geerat J. Vermeij

Department of Zoology, University of Maryland, College Park, Maryland 20742, USA

The phenotypic stability of many species in the face of changing conditions suggests that adaptive evolution can occur only under limited circumstances. One of the necessary conditions may be the lack of genetic mixing between dispersed populations inhabiting different environments. Intertidal molluscs on the east coast of North America between Cape Cod and Nova Scotia were exposed to an increase in the abundance of shell breaking predators when the green crab *Carcinus maenas* spread gradually northward from Cape Cod in the first half of the twentieth century¹. The periwinkle *Littorina littorea*, which produces larvae that become widely dispersed, did not show an increase in shell thickness as an adaptation to shell-breaking predation². However, I show here that the dog whelk *Nucella lapillus*, which is poorly dispersed in the bottom-dwelling juvenile phase, did adapt phenotypically after establishment of the green crab. This suggests that phenotypic stasis and gradual change are alternative responses depending on the degree of genetic mixing between populations.

Experiments show that *N. lapillus* co-occurring with predaceous crabs have slender, heavy, strong shells with a small aperture and a thick adult outer lip, whereas individuals from sites without crabs have thinner, weaker, stouter shells with a broader aperture and a thinner adult lip^{2–4}. Comparisons among populations further show that the degree of development of breakage-resistant traits is greatest where the abundance and strength of predators are greatest^{3,4}. Although shell shape is influenced by food availability and is size-dependent^{5–7}, it is at least in part determined genetically⁸. If *N. lapillus* was affected by the arrival of the green crab and if it responded adaptively to this change, spire height (shell height divided by aperture height) and mean adult lip thickness would be expected to increase through time.

Proof that the establishment of the green crab north of Cape Cod represented an increase in the importance of breakage as a component of selection in *N. lapillus* comes from data on the incidence of repaired shell injuries. The presence of a repaired injury provides evidence that the gastropod encountered a shell-breaking agent and that the shell was sufficient to withstand the agent's attack^{9,10}. The frequency of repaired injury (number of scars per shell) is probably an underestimate of the incidence of unsuccessful attack by shell-breaking agents in *N. lapillus*, because the thick lip of the adult may remain undamaged after an attack, so that no record of the encounter is preserved¹⁰. An increase in the frequency of repaired injury through time would mean that a greater proportion of individuals was exposed to potential selection in favour of resistance to breakage.

I examined 44 samples of *N. lapillus* (US National Museum of Natural History, Washington; Academy of Natural Sciences, Philadelphia; Museum of Comparative Zoology, Harvard University, Cambridge; and personal collections) from sites between Cape Cod and Halifax, the current northern limit of the green crab in North America¹¹. Samples were designated as pre-*Carcinus* or post-*Carcinus* depending on whether they were collected before or after the first known occurrence of the green crab at the locality in question^{12–14}. The 22 post-*Carcinus* samples were collected a minimum of 7 years after the arrival of the green crab. Scars resulting from repaired breaks of the outer lip were counted on the last whorl of each shell. The frequency of scars (Table 1) rose from 0.025 ± 0.021 in pre-*Carcinus* samples to 0.044 ± 0.028 in post-*Carcinus* samples ($P < 0.006$, one-tailed Mann-Whitney *U*-test). Scars increased in frequency at each of six sites where both pre-*Carcinus* and post-*Carcinus* samples were available.

N. lapillus evidently responded adaptively to the arrival of the green crab north of Cape Cod. Spire height increased from 1.385 ± 0.039 to 1.412 ± 0.061 ($P < 0.02$), and mean adult lip thickness increased from 1.92 ± 0.54 to 2.15 ± 0.49 ($P < 0.05$) after the appearance of the green crab (Table 1). At sites where both pre-*Carcinus* and post-*Carcinus* samples were available, spire height and lip thickness either increased (three sites) or remained about the same (three sites). These changes were not associated with a general increase in body size, and occurred despite the probable action of such other selective agents as wave surge^{2,15}, visual predators^{16–18} and parasites¹⁹.

Table 1 Differences in *N. lapillus* collected before and after the arrival of the green crab

Locality and date	n	Frequency of repaired injury	Adult shell slenderness	Adult lip thickness
Pre-Carcinus samples				
Beverly Farms, Massachusetts, 1876	60	0.017	1.365	1.86
Harborsville, Nova Scotia, 1850s	27	0.037	1.334	2.33
Isle Au Haut, Maine, 1897	143	0.014	1.347	1.20
Eastport, Maine, 1872	41	0.024	1.390	2.01
Eastport, Maine, before 1918	33	0.060	1.405	1.24
St George, Maine, 1863	51	0.020	1.385	2.56
Island near Rockland, Maine, 1912	59	0.018	1.411	2.24
Kittery, Maine, 1877	56	0.036	1.380	2.84
Castine, Maine, 1878	50	0.020	1.359	1.20
Salem, Massachusetts, 1870s	64	0	1.472	2.16
Peak's Island, Maine, 1873	160	0.044	1.416	2.30
Ten Pound Island, Massachusetts, 1878	40	0.025	1.413	2.00
Newcastle, New Hampshire, 1877	36	0.028	1.368	1.60
Beverly Bridge, Massachusetts, 1870s	36	0.056	1.454	2.28
Grand Manan, New Brunswick, 1920	95	0.032	1.370	1.04
Blue Hill Bay, Maine, 1915	167	0.036	1.403	2.24
Gloucester, Massachusetts, before 1900	17	0	1.337	1.76
Mount Desert, Maine, around 1920	374	0.017	1.309	2.26
St Andrews, New Brunswick, 1920	157	0.038	1.369	0.94
Wells, Maine, 1915	19	0.053	1.375	2.62
Islesboro, Maine, 1894	17	0	1.423	1.50
Somes Sound, Maine, around 1920	373	0.011	1.390	2.44
Post-Carcinus samples				
Kennebunkport, Maine, 1925	44	0.070	1.392	2.32
Cape Neddick, York, Maine, 1951	74	0.066	1.320	1.44
St Andrews, New Brunswick, 1963	79	0.051	1.408	1.44
Kittery, Maine, 1951	85	0.048	1.449	2.86
Biddeford, Maine, 1938	36	0.083	1.400	2.28
Squirrel Island, Maine, 1971a	13	0	1.344	1.21
Squirrel Island, Maine, 1971b	11	0.091	1.527	2.50
Harbourville, Nova Scotia, 1960	40	0.075	1.513	2.86
Lincolntonville, Maine, 1941	62	0.016	1.424	1.46
North Weymouth, Massachusetts, 1914	21	0.14	1.473	2.44
Salisbury Bay, New Brunswick, 1957	21	0.048	1.475	2.26
Gloucester, Massachusetts, 1925	107	0.036	1.324	2.54
Queensland, Nova Scotia, 1960	68	0.044	1.499	2.00
Marblehead, Massachusetts, 1927	78	0.013	1.415	2.70
Lynn, Massachusetts, 1938	15	0.067	1.456	1.56
Lawrencetown, Nova Scotia, 1960	44	0	1.312	2.36
Kelly's Cove, Nova Scotia, 1959	68	0.044	1.396	2.32
Alma, New Brunswick, 1960	31	0	1.422	2.50
Ogunquit, Maine, 1941	53	0.038	1.435	1.62
Eastport, Maine, 1981	50	0.040	1.399	1.72
Nahant, Massachusetts, 1981	65	0.091	1.431	2.90

n, Number of shells in the sample.

the phenotypes of species as major changes in the biological environment are being brought about by man.

Received 11 May; accepted 3 August 1982.

1. Vermeij, G. J. *Evolution* **36**, 561-580 (1982).
2. Kitching, J. A., Muntz, L. & Ebling, F. J. *J. Anim. Ecol.* **35**, 113-126 (1966).
3. Hughes, R. N. & Elner, R. W. *J. Anim. Ecol.* **48**, 65-78 (1979).
4. Currey, J. D. & Hughes, R. N. *J. Anim. Ecol.* **51**, 47-56 (1982).
5. Moore, H. B. *J. mar. Biol. Ass. U.K.* **21**, 61-89 (1936).
6. Spight, T. M. *J. exp. mar. Biol. Ecol.* **13**, 215-228 (1973).
7. Phillips, B. F., Campbell, N. A. & Wilson, B. R. *J. exp. mar. Biol. Ecol.* **11**, 27-69 (1973).
8. Lagen, M. D. *Proc. Malac. Soc.* **39**, 383-388 (1971).
9. Vermeij, G. J. *Malacologia* **23**, 1-12 (1982).
10. Vermeij, G. J., Schindel, D. E. & Zipser, E. *Science* **214**, 1024-1026 (1981).
11. Bousfield, E. L. & Laubitz, D. R. *Nat. Mus. Canada Publ. Biol. Oceanogr.* **5**, 1-51 (1972).
12. Dow, R. L. & Wallace, D. E. *Fish. Circ.* **8**, Bull. Dept Sea Shore Fish., Augusta, Maine, 11-15 (1952).
13. Scattergood, L. W. *Fish. Circ.* **8**, Bull. Dept Sea Shore Fish., Augusta, Maine, 2-10 (1952).
14. Welch, W. R. *Fish. Bull.* **67**, 337-345 (1968).
15. Berry, R. & Crothers, J. H. *J. Zool.* **155**, 5-17 (1968).
16. Colton, H. S. *Ecology* **3**, 146-157 (1922).
17. Berry, R. J. & Crothers, J. H. *J. Zool.* **162**, 293-302 (1970).
18. Berry, R. J. & Crothers, J. H. *J. Zool.* **174**, 123-148 (1974).
19. Feare, C. J. *Bird Study* **18**, 121-129 (1971).
20. Scheltema, R. S. *Marine Organisms* (eds Battaglia, B. & Beardmore, J. A.) 303-322 (Plenum, New York, 1978).

A molecular explanation of frequency-dependent selection in *Drosophila*

Yousef Haj-Ahmad

Department of Biological Sciences, Brock University, St Catharines, Ontario, Canada L2S 3A1

Donal A. Hickey

Department of Biology, University of Ottawa, Ottawa, Ontario, Canada K1N 6N5

Frequency-dependent selection provides a means for maintaining genetic variability within populations, without incurring a large genetic load^{1,2}. There is a wealth of experimental evidence for the existence of frequency-dependent changes in genotypic fitness among a wide variety of organisms. Examples of traits which have been shown to be subject to frequency-dependent selection include the self-incompatibility alleles of plants³, chromosomal rearrangements in *Drosophila*⁴, visible mutations⁵, enzyme variants⁶⁻¹¹ and rare-male mating advantage in *Drosophila*¹²⁻¹⁴. These experiments have been interpreted in a number of different ways¹⁵. Principally, frequency dependence of genotype fitness may result from intergenotype facilitation due to the production of biotic residues¹⁶⁻²⁰, or from the differential use of resources by the competing genotypes²¹. However, it has proved extremely difficult to isolate and identify any biotic residue of importance or, alternatively, to understand the manner in which genotypes partition the environment. Thus, the difficulty in the interpretation of experiments which show frequency-dependent selective effects stems largely from our lack of understanding of the exact physiological mechanisms which produce these frequency-dependent effects². The principal aim of this study was to investigate the mechanisms associated with frequency-dependent selection at the amylase locus in *Drosophila melanogaster*. The excretion of catalytically active amylase enzyme and its effects on food medium composition were correlated with the outcome of intraspecific competition between amylase-deficient and amylase-producing genotypes. Amylase-producing genotypes were shown to excrete enzymatically active amylase protein into the food medium. The excreted amylase causes the external digestion of dietary starch; this accounts for the frequency-dependent increase in the viability of the amylase-deficient mutants in mixed cultures, maintained on a starch-rich diet.

The results described here involved two homozygous amylase genotypes, one which produces active amylase enzyme, *Amy*¹ (refs 22, 23); and another which has no detectable amylase activity, *Amy*^{null} (Haus and Hickey, unpublished data). Our

Insufficient material of *N. lapillus* was available to test whether shell shape and frequencies of repair remained constant over time in Canada north of the range of the green crab. For *L. littorea*, however, both shell thickness and the frequency of repair have remained the same over at least the last 100 years¹. Although the arrival of the green crab was associated with an increase in the frequency of repair in both *L. littorea* and *N. lapillus*, only the latter species showed signs of adaptive phenotypic change. These results suggest that genetic mixing was in part responsible for phenotypic stasis, and support the view²⁰ that limited dispersal permits relatively rapid evolutionary change. They also point to the importance of monitoring

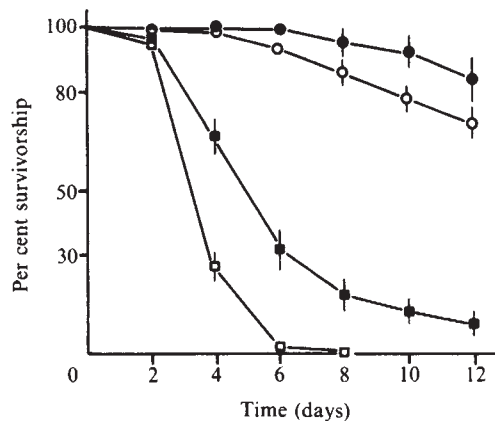


Fig. 1 Relationship between initial frequency and per cent survivorship of *Amy*^{null} flies, when grown in mixed culture with individuals of the *Amy*¹ genotype. Culture vials (8 dram shell vials) contained 15 ml of a starch-rich medium (8% soluble starch+0.5% killed brewer's yeast). In mixed cultures, genotypes were usually identified by an association with visible genetic markers (curved wings and Lobe eyes). Experiments were repeated using reversed marker associations, to eliminate the possibility of marker effects. Initially, a total of 100 one-week-old flies were added to each vial. Dead flies were removed every 24 h. Survivors were transferred to fresh vials every 48 h. Initial *Amy*^{null} frequencies were: 1.0 (□), 0.9 (■), 0.6 (○) and 0.1 (●). These data represent the means of seven replicate experiments. Bars indicate s.e.m.

prediction was that the *Amy*^{null} flies would be unable to use dietary starch as an energy source. This prediction was tested by measuring the viability of adult flies of both genotypes at 2-day intervals over a 12-day period. Three different dietary regimes were used: one food medium containing 8% starch; a second medium containing 8% glucose in place of starch; and a third medium containing no carbohydrate. Table 1 shows the per cent survivorship recorded on day 12 of the experiment. The survivorship of both genotypes on the food which contained glucose was over 95%; the survivorship of both strains in the absence of all carbohydrates was less than 1%. This indicates that the dietary sugar is essential for the survival of both genotypes. When dietary sugar was replaced by starch, the survivorship of the *Amy*¹ homozygotes remained high (over 90%) but the survivorship of the *Amy*^{null} genotype fell to zero (Table 1). Thus, as expected, the *Amy*¹ flies can use dietary starch as an energy source, but the *Amy*^{null} individuals are unable to do so.

The next set of experiments followed the same general procedure as outlined above, but the two genotypes were main-

Table 1 Per cent survivorship of two amylase genotypes on three food media

Genotype	Food type		
	Glucose (8%)	No carbohydrate	Starch (8%)
<i>Amy</i> ¹	96.0 ± 1.1	0.2 ± 0.1	92.1 ± 1.9
<i>Amy</i> ^{null}	96.3 ± 1.3	0	0

All foods contained 0.5% killed brewer's yeast and 0.8% propionic acid (as a mould inhibitor), and 1.5% agar. In these conditions of dietary deprivation (low yeast) the supplementary carbohydrates were essential for survival. In contrast, in the presence of high levels of yeast (5%), the survivorship of both strains in the absence of supplementary carbohydrates was quite high. Before the experiments were begun, all flies were aged for one week on a nutritionally rich diet (5% killed yeast, 10% sucrose). All cultures were maintained in 8 dram shell vials at 25 °C. Experiments were initiated with 100 one-week-old flies of each genotype. Dead flies were removed every 24 h and survivors were transferred to a fresh medium every 48 h. Data represent survivorship on day 12 of the experiment. Each value represents the mean of 10 replicate experiments ± s.e.m.

tained in mixed rather than pure cultures. Figure 1 shows that the survivorship of *Amy*^{null} on starch food in mixed culture was greater than in pure cultures, with survivorship increasing as the initial frequency of the *Amy*^{null} genotype decreased; that is, the fitness of the *Amy*^{null} genotype in mixed cultures shows a pattern of negative frequency dependence. However, at low frequency, the *Amy*^{null} survivorship only approaches that of *Amy*¹, but does not surpass it. The survivorship of *Amy*¹ flies in mixed culture was high, and comparable with that seen in pure cultures, regardless of the initial genotype frequencies; therefore, one would not expect that this particular example of frequency-dependent selection would result in a stable genetic polymorphism. The high survivorship of the *Amy*¹ flies in all conditions indicates that the general level of competition must be rather low.

As the survival of the *Amy*^{null} genotype depends on the availability of dietary sugar (Table 1), it seemed likely that sugar was being made available to these flies in mixed cultures fed dietary starch. This hypothesis was tested directly by assaying the free sugar concentration in a food medium containing 8% added starch, after it had been exposed to varying numbers of amylase-producing flies. Figure 2 shows that as the number of *Amy*¹ flies was increased, the sugar content of the food medium also increased. Therefore, when the frequency of the *Amy*^{null} genotype was low in mixed cultures and the frequency of the *Amy*¹ genotype was correspondingly high, the sugar content of the starch containing medium was elevated. This explains the frequency-dependent survival of the *Amy*^{null} flies in mixed cultures (Fig. 1).

Amylase-producing flies may modify their environment by either direct excretion of sugars or excretion of active amylase which, in turn, causes the external digestion of starch. Amylase activity was measured at the surface of a medium exposed to *Amy*¹ males for 48 h. A significant amount of amylase activity (about 20% of that recorded in the flies themselves) was detected. We cannot deduce the actual rate of amylase excretion, however, as we do not know the stability of the enzyme over the 48-h period. There was no detectable increase in the amylase activity of food media which were exposed to *Amy*^{null} flies.

Finally, we checked that the external amylase activity that was associated with amylase-producing genotypes was indeed *Drosophila* amylase, and not due to microbial contamination.

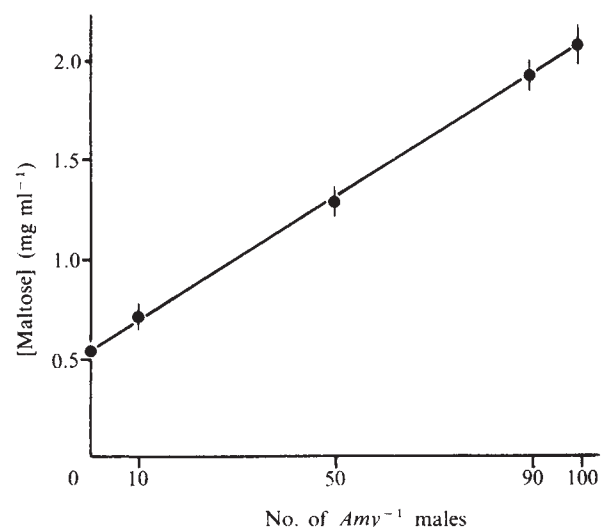


Fig. 2 Relationship between sugar concentration in the food medium and exposure to amylase producing flies (genotype *Amy*¹). The food medium contained 8% soluble starch and 0.5% killed brewer's yeast. Varying numbers of flies (males only) were maintained on the medium for 48 h. Sugar concentration is expressed in maltose equivalents, as measured by the DNSA assay²⁴. Values given are the means of 12 replicate experiments. Bars indicate s.e.m.

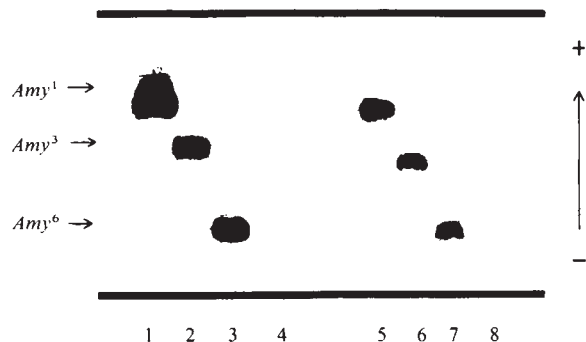


Fig. 3 Comparison of electrophoretic patterns for four amylase genotypes and the enzymes which were recovered from their respective substrates. One hundred males (one week old) were maintained in clean glass vials for 10 h. At the end of this period flies were homogenized and the excretion recovered in 0.4 ml of electrophoresis buffer. Samples of both the crude fly homogenates and their respective excretions were electrophoresed in vertical polyacrylamide gels. Gels were incubated in a starch solution and stained with iodine-potassium-iodide to visualize amylase activity²³. Electromorph designations are indicated to the right of the figure. Gel samples 1 to 4 are homogenates of *Amy*¹, *Amy*³, *Amy*⁶ and *Amy*^{null} flies, and samples 5 to 8 are their respective excretions. This is a negative print of a gel photograph.

This was done by electrophoresis of a number of electrophoretically distinct amylase variants; flies of each genotype were homogenized and electrophoresed along with a sample of the associated food medium. It can be seen that, in each case, the amylase which was recovered from the substrate was electrophoretically identical to that of the flies maintained on this substrate (Fig. 3). Therefore, it seems that the modification of the starch-containing medium by the amylase-producing flies is due to the excretion of active enzyme molecules. Thus, the enzymatic products of one genotype positively enhances the survivorship of the other genotype which lacks this enzyme.

In summary, frequency-dependent selection is a potentially powerful mechanism for the maintenance of genetic polymorphism^{1,2}. However, as pointed out by Clarke², despite the success of many researchers in detecting frequency-dependent selection and the power of that selective mode in maintaining genetic variations, the mechanisms involved are still largely unknown. The results presented here provide a unique example of a noncompetitive intergenotype interaction, one which has a frequency-dependent effect on the viability of one of the genotypes and one where we fully understand the underlying physiological mechanisms. However, the particular form of frequency-dependent selection discussed here will not necessarily maintain a polymorphism, because the fitness of the null allele is always less than, or equal to, the fitness of the wild-type amylase-producing allele.

We thank Dr Winifred Doane for providing the *Amy*¹ stock. Financial support was provided by a grant from NSERC, Canada. Y.H.-A. was the recipient of an Ontario Graduate Scholarship.

Received 5 May; accepted 13 July 1982.

- Lewontin, R. C. *The Genetic Basis of Evolutionary Change* 256-260 (Columbia University Press, New York, 1974).
- Clarke, B. C. *Proc. R. Soc. B* **205**, 453-474 (1979).
- Mather, K. *Genetical Structure of Populations* (Chapman and Hall, London, 1973).
- Levene, H., Pavlovsky, O. & Dobzhansky, T. *Evolution* **8**, 335-349 (1954).
- Clarke, B. *Heredity* **24**, 347-352 (1969).
- Kojima, K. & Huang, S. L. *Evolution* **26**, 313-321 (1972).
- Kojima, K. & Tobari, Y. N. *Genetics* **61**, 201-209 (1969).
- Nassar, R. *Genetics* **91**, 327-338 (1979).
- Snyder, M. & Ayala, F. J. *Genetics* **92**, 995-1003 (1979).
- Tosic, M. & Ayala, F. J. *Genetics* **97**, 679-701 (1981).
- Morgan, P. *Nature* **263**, 765-766 (1976).
- Petit, C. *Bull. Biol. Fr. Belg.* **92**, 248-329 (1958).
- Ehrman, L. *Am. Nat.* **101**, 415-424 (1967).
- Spiess, E. B. *Am. Nat.* **102**, 363-379 (1968).
- Ayala, F. J. & Campbell, C. A. *A. Rev. ecol. Syst.* **5**, 115-138 (1974).

- Lewontin, R. C. *Evolution* **9**, 27-41 (1955).
- Ayala, F. J. *Science* **171**, 820-824 (1971).
- Dolan, M. & Robertson, A. *Heredity* **35**, 311-316 (1975).
- Bos, M. *Evolution* **33**, 768-771 (1979).
- Caligari, P. D. S. *Heredity* **45**, 219-231 (1980).
- Ayala, F. J. *Am. Sci.* **60**, 348-357 (1972).
- Doane, W. W. *J. exp. Zool.* **171**, 321-341 (1969).
- Hickey, D. A. *Biochem. Genet.* **19**, 783-793 (1981).
- Noelting, G. & Bernfield, P. *Helv. chim. Acta* **31**, 286-290 (1948).

Analysis of discrimination mechanisms in the mammalian olfactory system using a model nose

Krishna Persaud & George Dodd

Warwick Olfaction Research Group, Department of Chemistry and Molecular Sciences, University of Warwick, Coventry CV4 7AL, UK

Olfaction exhibits both high sensitivity to odours and high discrimination between them¹. We suggest that to make fine discriminations between complex odorant mixtures containing varying ratios of odorants without the necessity for highly specialized peripheral receptors, the olfactory systems makes use of feature detection using broadly tuned receptor cells organized in a convergent neurone pathway. As a test of this hypothesis we have constructed an electronic nose using semiconductor transducers and incorporating design features suggested by our proposal. We report here that this device can reproducibly discriminate between a wide variety of odours, and its properties show that discrimination in an olfactory system could be achieved without the use of highly specific receptors.

The morphology of the olfactory system in mammals is remarkably homogeneous² (Fig. 1). Typically, the sensing mucosa has some tens of millions of sensing cells, which act as the primary neurones and synapse with the secondary neurones in the olfactory bulb. The receptor cells in all mammals show a small range of variation, are broadly tuned and respond to a wide variety of odorant types². Cells responding to a single class of odours, for example musks, have not been found. The molecular transducers have not been identified; the sensory

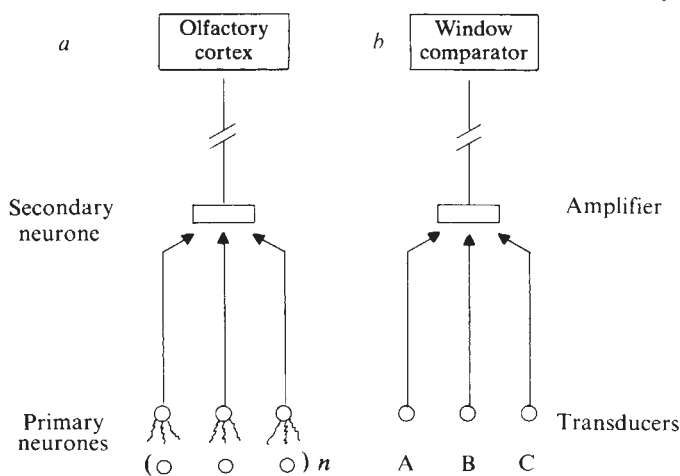


Fig. 1 Diagrammatic comparison of the mammalian olfactory system and the model nose. *a*, Olfactory system. The sensing cells are the ciliated primary neurones. The axons of these neurones pass back through the cribriform plate into the olfactory bulb where they synapse with the secondary neurones, the mitral cells. The signal from the mitral cells passes through the higher olfactory pathways into the olfactory cortex. There is a striking convergence in the peripheral pathway; the ratio of primary to secondary neurones is of the order of $2 \times 10^4:1$. *b*, Model nose. The sensing elements are semiconductor devices with overlapping odorant response distributions. The amplifier, an analogue of the secondary neurone, measures the ratio of the response of selected transducers using a defined algorithm. The signal passes to the window comparator and associated memory circuits. A particular type of odour activates a designated light-emitting diode.

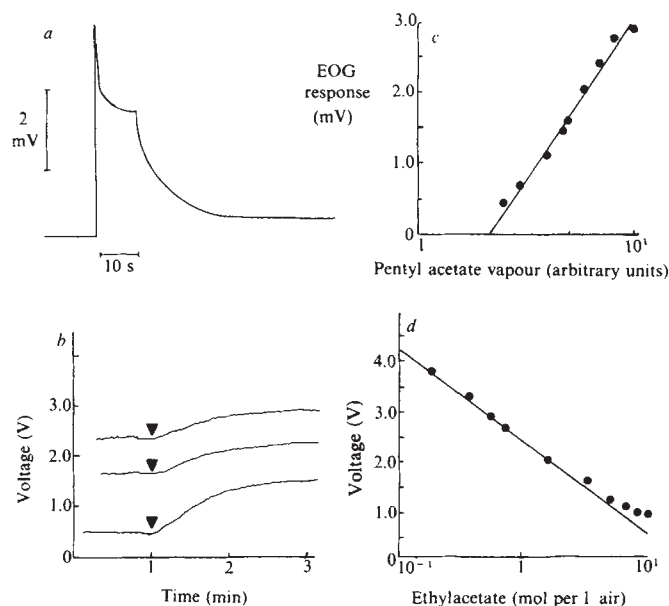


Fig. 2 Comparison of the response characteristics of the olfactory system and model nose. *a*, The summated receptor potential (electro-olfactogram) in olfactory mucosa to a square-wave pulse of pentyl acetate. The electro-olfactogram was recorded as described previously²¹. *b*, The voltage change of a Figaro 812 semiconductor gas sensor in response to the odorant ethyl trimethylacetate. The sensor was located at the top of a 5 l flask immersed in a water bath at 25 °C. The humidity in the flask was maintained at 100% and the air in the flask was mixed using a paddle stirrer. The odorant, 0.1 µl, was injected into the flask through a gic septum using a Hamilton syringe and vaporized in these conditions. Time of injection is marked ▼. *c*, The electro-olfactogram from sheep olfactory mucosa as a function of odorant concentration. The odorant was pentyl acetate. The odorant was delivered using a precision olfactometer and the odorant concentration was determined using a flame ionization detector as previously described^{21,22}. Ten-second odour pulses with a pulse interval of 2 min were used. In these conditions, there was no adaptation of the response. An ascending concentration series was used to avoid adaptation. *d*, The voltage response of the Figaro 812 sensor as a function of odorant concentration. The response was measured as described in *b*, using an ascending concentration series. Sufficient time was allowed for the response to reach the plateau region (see *b*).

membranes contain many particles which may be receptor proteins³. On average, 10–20,000 primary neurones synapse with a single secondary neurone. The secondary neurones show a fairly specific response to odour classes but the response specificities of the primary neurones connected to any particular secondary neurone are unknown. There are other neurone types in the olfactory bulb which cross-connect the secondary neurones.

It is clear that the discrimination properties of the olfactory system are a property of the system as a whole. The higher centres process the pattern of signals from lower centres and the pattern is used to encode a particular type of odour. It should be possible to stimulate this process using some of the principles of pattern classification which have come from studies on artificial intelligence^{4–6}. Extensive mathematical modelling of pattern recognition systems has been carried out^{4,5} and a specific model for the recognition of chemical stimuli has been formulated⁷.

To construct an electronic model olfactory system, we have selected two fundamental characteristics of all mammalian olfactory systems: (1) the odorant detectors, the primary neurones, respond to a wide range of chemical types; and (2) the output from the detectors is processed in a convergent feature detection system which has, at least in part, parallel feature detection. Thus, we have made the following assumptions for the design of an electronic nose which can learn to discriminate between smells: there is no requirement for odour-specific transducers; and the ratio of the signals from the transducers can be processed to identify an odour.

The electrical response of the primary neurones (R) is proportional to the logarithm of the odorant concentration, $[O]$, over a range of concentrations (Fig. 2c) and can be described by the equation

$$R = \alpha \log [O] \quad (1)$$

where α is the odorant constant and is characteristic of a particular odorant. Because the neurones are generalist detectors in a parallel detection system, the generator potential of the receptor cells must contain information about both the quality and intensity of an odour and this information must be encoded in the odour constant, α . The information on odour quality can be separated from that on odour intensity by an analysis of the ratios of the responses from the transducers. The simple method we have selected for the nose is described in Fig. 3. This shows the hypothetical frequency distribution of the responses of a sensing cell to different odours. Such a distribution has not been measured, both because of technical difficulties arising in recording from the small olfactory neurones and because of the lack of an appropriate physico-chemical parameter for the abscissa in this type of plot. The actual distributions may be non-gaussian and multi-modal. The odorants giving the maximum response for a cell may have the optimum fit to the binding sites on the sensory membranes and such odorants may be regarded as 'primary' odorants.

In the example shown (Fig. 3), the two transducers have overlapping response distributions and respond differentially to the two odorants. The ratio of the responses of each transducer (T) towards the two odorants is given by the expression

$$\frac{T_1}{T_2} = \frac{R_1/R_2}{R_4/R_3} = \frac{\frac{\alpha_1 \log [O]_A}{\alpha_2 \log [O]_B}}{\frac{\alpha_4 \log [O]_A}{\alpha_3 \log [O]_B}} = \beta \quad (2)$$

Table 1 Response of the artificial nose to a wide variety of odours

α , Odorant	Selectivity constants using the following pairs of transducers:		
	812/813	812/817	711/813
Isojasmone	0.12	1.58	—
Undecanaldehyde	0.13	1.85	—
Methanol	0.14	1.60	—
Clove bud oil	0.18	1.68	—
Rose oil (synthetic)	0.23	1.49	—
Phenylethanol	0.33	0.88	—
Acetophenone	0.40	1.90	—
Cineole	0.42	1.24	—
Lilac (synthetic)	0.48	—	—
Benzyl methyl ketone	0.71	—	—
Coriander oil	1.43	—	—
Amyl alcohol	1.51	2.43	—
Amyl acetate	1.73	—	1.25
Jasmin oil (synthetic)	1.80	1.95	—
Clary sage oil	1.89	—	—
Isobutyl acetate	1.95	2.33	1.81
Verival	1.96	1.90	—
Ethanol	2.14	2.33	1.98
Diethyl ether	2.34	2.33	2.27
Pulegone	—	—	1.06
2-Acetyl-3-methylpyrazine	—	0.72	0.11

b, Relationship between sensor response and odorant concentration for different odours for a single transducer

Odorant	$-\alpha$
Butyl propionate	1.54
Ethyl acetate	1.64
Pentyl acetate	1.65
Ethyl trimethylacetate	1.92
Ethyl bromoacetate	1.94
Ethyl propionate	1.95
Cineole	2.06

The reproducibility of the measurements in *a* can be judged from the data in Fig. 4. The selectivity constant (β) is defined in equation (2). The voltage response of a Figaro 812 gas sensor was measured as a function of the odorant concentrations as described in Fig. 2 legend. A plot of the response against the log odorant concentration was linear (Fig. 2). The slope of this line gives the odour constant, α .

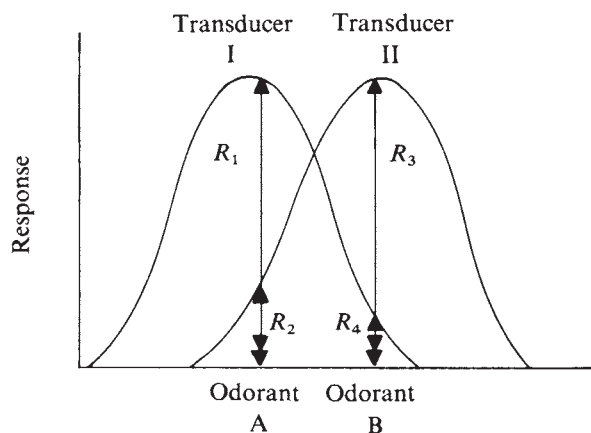


Fig. 3 The odour frequency response for a fixed concentration of odour of two hypothetical odour transducers. The odorants A and B respectively elicit responses R_1 and R_4 from the first transducer and responses R_2 and R_3 from the second transducer.

if the transducers respond to the odorants like the olfactory neurones (equation (1)). The selectivity constant, β , is independent of the odorant concentration and can be used as an index to describe an odour. This analysis is easily extended to N transducers with overlapping frequency distributions and the ratio of the responses for a particular odour computed with the appropriate algorithm will occupy a unique point in a multi-dimensional space. This extension is necessary in the case of the real nose where there are tens of millions of primary neurones and each secondary neurone, on average, is synapsed to about 2×10^4 sensing cells. The model nose can explore the discrimination by following these principles and using the minimum number of transducers (two).

Several odour-detector systems have previously been described⁸⁻¹⁴, but they have a limited application. More recently, n -type semiconductor devices have been described which show a conductance change when they adsorb combustible gases^{15,16}. An important feature of the semiconductor devices is that the response spectrum to vapours can be slightly altered by doping with suitable materials.

Odour sensors based on copper-doped tin oxide semiconductors were constructed exactly as described by Ohno¹⁵. These devices responded to a wide variety of odours. The relative responses of one particular device to ethanol, ammonia, diethyl-ether and iodomethane were respectively 1.0, 0.17, 0.07 and 0.05.

The commercially available semiconductor gas detectors are more stable than the devices we constructed and were used as the transducers for the nose. These transducers, like the sensory neurones, gave a voltage change proportional to the log of the odorant concentration (Fig. 2a), but the response time (Fig. 2a, b) was considerably slower than the response time of the olfactory system. The transducers, in contrast to the neurones, decrease their voltage in response to an odorant stimulus; this difference does not affect the pattern recognition mechanism. A single type of transducer responds differentially to different odours and this is reflected in the odour constants shown in Table 1. Over a period of 6 months the transducers gave a consistent response with most odorants tested and duplicate measurements agreed to within $\pm 5\%$. Some odorants, including 1,8-cineole, 5- α -androstane-3-one and pyridine, produced irreversible changes in the detector and were excluded from the main experiments.

Three Figaro gas sensors were used as the transducers: the 812(A), a general purpose combustible gas sensor; the 813(B), which is more sensitive for alcohols; and the 711(C), which is more sensitive for carbon monoxide. The relative sensitivities of these transducers towards ethanol are 1.0, 0.4 and 10.0 respectively. Thus, they fulfill our requirement of a differential response spectrum. The transducers were arranged in a circuit so that the ratio of their outputs could be derived (Fig. 1). The

precision of the measurements was satisfied by a simple analog circuit¹⁷ which was modelled on a published circuit¹⁸.

The device responds to a wide variety of odours, but since it has no odour memory it will not identify the odour. Thus the nose has to be calibrated by exposing it to various odours and locating the response in the memory. The human olfactory system appears to have similar properties. The selectivity constants determined for a variety of odorants cover a range of values (Table 1a) and allow the nose to discriminate between a large variety of odours. This discrimination is indicated in our system by the operation of a numbered light-emitting diode. The nose reliably discriminates between odours (Fig. 4).

Using three transducers, the system can mimic the discrimination of the mammalian olfactory system at a gross level. An electronic nose which can make fine discriminations, and which will be of interest as a quality control device in industries concerned with flavours, perfumes and smells is feasible using the principles we have used for the simple nose, together with recent developments in microelectronics¹⁹.

The implications of the model nose for olfactory receptor mechanisms are interesting. The model shows that discrimination between odours may be effected by the olfactory system amplifying the small differences in the responses from the detector cells. These small differences in the responses of the primary cells may arise from relatively nonspecific interaction of the odorants with the sensory membranes. Specific receptor proteins such as are found for many hormones and neurotransmitters and some odorants^{3,14,20} may exist only for odorants of special biological importance such as sexual pheromones.

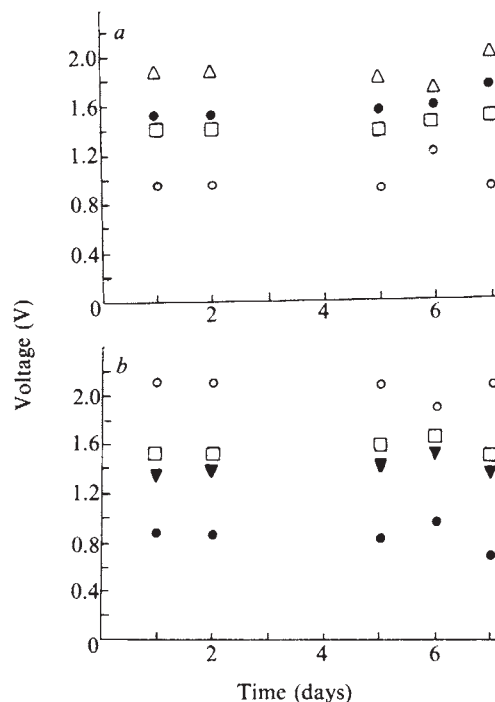


Fig. 4 Reproducibility of the response of the model nose to selected odorants. The odorants had significantly different selectivity constants (Table 1). Response bands were selected on the voltage comparator with limits of $\pm 10\%$ of the appropriate ratio (selectivity constant). a, Responses measured by taking the ratios of the outputs from sensors A (812) and B (813). Δ , Amyl acetate; $\bullet$, ethanol; $\square$, ether; $\circ$, octanol. b, Responses measured by taking the ratios of the outputs from sensors A (812) and C (711). $\circ$, Valeric acid; $\square$, methylpyrazine; ∇ , lemon oil; $\bullet$, isojaesmane. Each transducer had its own zero circuitry which allowed us to compensate the response for changes in humidity or background odours and which converted the drop in resistance of the transducer to a positive voltage increase. The voltages were fed to individual temperature-stabilized logarithmic amplifiers. The ratio of the signals from the transducers was derived by subtracting the outputs of the logarithmic amplifiers using a differential amplifier. The ratioed signal, after passage through an anti-logarithmic amplifier, was fed into a window comparator. In the experiments reported here we measured the ratio of the signals from transducers A and B, and A and C respectively (Fig. 1).

The work of the Warwick Olfactory Research Group is supported by grants from the SERC, MRC and the Royal Society. We thank Dr H. Mykura for help with the construction of the semiconductor transducers.

Received 24 March; accepted 13 July 1982.

1. Stoddart, D. M. (ed.) *Symp. zool. Soc., Lond.* **45**, (1980).
2. Dodd, G. H. & Squirrel, D. J. *Symp. zool. Soc., Lond.* **45**, 35–56 (1980).
3. Dodd, G. H. & Persaud, K. in *Biochemistry of Taste and Olfaction* (eds Cagan, R. & Kare, M.) 333–357 (Academic, New York, 1981).
4. Holden, A. V. *Lect. Notes Biomath.* **32** (Springer, Heidelberg, 1976).
5. Sampson, J. R. *Adaptive Information Processing* (Springer, Berlin, 1976).
6. Uttal, W. R. *The Psychobiology of Sensory Coding* (Harper & Row, New York, 1973).
7. Deutsch, S. *Models of the Nervous System* (Wiley, New York, 1967).
8. Cherry, R. J., Dodd, G. H. & Chapman, D. *Biochim. biophys. Acta* **211**, 409–416 (1970).
9. Wilkens, W. F., Hartman, J. D. & Rosano, H. *Ann. N.Y. Acad. Sci.* **116**, 608–612 (1964).
10. Moncrieff, R. W. J. *appl. Physiol.* **16**, 742–749 (1961).
11. Dravnieks, A. & Trotter, P. J. *J. Sci. Instrum.* **42**, 624–627 (1965).
12. Sawateri, K. *Chem. Abstr.* **83**, P 35757 T (1975).
13. Tanyolac, N. N. & Eaton, J. R. *J. Am. pharmac. Ass.* **39**, 10–14 (1950).
14. Persaud, K., Pelosi, P. & Dodd, G. H. *Olfaction & Taste* **7**, 101 (1980).
15. Ohno, Y. *Chem. Abstr.* **85**, 28276Y (1976).
16. T.G.S. Gas-Sensing Semi-Conductor (Figaro Engineering Inc., Osaka, 1974).
17. Dodd, G. H. & Persaud, K. C. *Provisional UK Patent No.* 8035806 (1980).
18. *Electronics Today International* 62–65 (January 1979).
19. Roberts, G. G., Pande, K. P. & Barlow, W. A. *Electron. Lett.* **13**, 581–583 (1977).
20. Persaud, K., Wood, P., Squirrel, D. J. & Dodd, G. H. *Biochem. Soc. Trans.* **9**, 107–108 (1981).
21. Meneve, A., Dodd, G. H. & Poynder, T. M. *Biochem. J.* **176**, 845–854 (1978).
22. Squirrel, D. J. thesis, Univ. Warwick, UK (1978).

Corticotropin releasing activity of the new CRF is potentiated several times by vasopressin

G. E. Gillies, E. A. Linton & P. J. Lowry

Pituitary Hormone Laboratory, Department of Chemical Pathology, St Bartholomew's Hospital, London EC1A 7BE, UK

Initially the hypothalamic factor responsible for the release of corticotropin (CRF), was thought to be a simple peptide^{1–6}. More recent work has led to the conclusion that CRF is a multifactorial complex^{7–9}. In 1979 we proposed¹⁰ that vasopressin, much disputed as a CRF candidate, was a major constituent of the complex, interacting with and potentiating the CRF activity of the other component(s). The recent characterization of a 41 residue ovine hypothalamic peptide capable of releasing adrenocorticotrophic hormone (ACTH) in a dose-related manner¹¹ has allowed us to compare its CRF bioactivity with that of vasopressin and simple extracts of the hypothalamus, and to investigate any interaction it may have with vasopressin and other hypothalamic factors in the release of ACTH. We report here that the new CRF is more potent than vasopressin in releasing ACTH. When given simultaneously with vasopressin a fourfold potentiation of CRF activity with steep dose-response characteristics was observed. It also potentiated vasopressin-free hypothalamic extracts, suggesting that the new CRF does not account for all the non-vasopressin portion of the CRF complex.

The bioassay used in this study was that described previously¹², except that the antioxidant, ascorbic acid, was added to all physiological media (60 µg ml⁻¹) and peptide solutions under storage (1 mg ml⁻¹) for two reasons. First, we have observed synergism between chromatographically separated hypothalamic fractions only when ascorbic acid was included in the eluant¹⁰. Second, Vale *et al.*¹¹ observed that the synthetic 41 residue CRF was 10 times more potent than purified native ovine CRF. This was attributed to the presence of the methionine sulphoxide in the latter peptide and this oxidized form, which may have been spontaneously generated during purification, may not display all the biological characteristics of the native unoxidized form.

In agreement with Vale *et al.*¹¹, we observed that the new CRF was more potent than vasopressin in releasing ACTH (Fig. 1). However, the most active CRF in our system is still a rat stalk-median eminence extract (SME) which we have generally adopted as our CRF standard¹². A linear response to this extract begins around 0.01 SME ml⁻¹ (refs 12, 13) and at the

maximum concentration we have tested (0.2 SME ml⁻¹) ACTH release was at least 20-fold that of basal secretion (unpublished observations). In view of this large response, and for fear of nonspecific effects due to tissue extracts, our investigations have been routinely carried out in the linear range of responses observed for 0.01–0.1 SME ml⁻¹. We have never reached a plateau response with SME. Figure 1b illustrates that the response to CRF-41 plateaus by 60 ng ml⁻¹. The log dose-response curve to vasopressin is characteristically shallow and significantly non-parallel to that for SME with an ED₅₀ in the 5–10 ng ml⁻¹ range¹⁴. The highest dose of SME used here (0.1 SME ml⁻¹), however, contains only 1 ng of vasopressin by radioimmunoassay. Small-scale chromatography of rat SME has shown that at least three CRFs can be separated by gel filtration, one of which has the characteristics of vasopressin on chromatography, immunoassay and CRF bioassay¹⁰. The individual CRF peaks showed only a small fraction of the total CRF bioactivity, which was restored on recombination of the three partially purified CRF peaks, resulting in a log dose-response curve identical to that for crude SME. These results suggested that CRF consists of vasopressin plus at least two other factors which interact synergistically for full CRF bioactivity. We therefore tested synthetic vasopressin and CRF-41 simultaneously in our bioassay and observed a dramatic potentiation (Fig. 1b). A threefold elevation of ACTH release above background when either peptide was given alone became an 11-fold elevation at a dose of 10 ng ml⁻¹ of each peptide, and the resulting log dose-response curve was as steep as that for crude SME. Thus the combination of vasopressin and the new CRF clearly constitutes a potent stimulator of ACTH release from the anterior pituitary corticotropes in our assay system.

Next we investigated whether the CRF activity of SME was simply due to vasopressin plus the new CRF or whether other factors are involved. When the new CRF was added to rat SME a considerable potentiation of CRF bioactivity was observed with the higher doses of CRF tested (Fig. 2). This suggests that it is synergizing with vasopressin present in the extract and/or other factors. To investigate this interaction further we tested the effects of adding the new CRF to SME which has its ACTH-releasing activity quenched by vasopressin antiserum (Fig. 2). The potentiating ability of the new CRF at 10 ng ml⁻¹ was reduced by almost one-half, with the observed responses

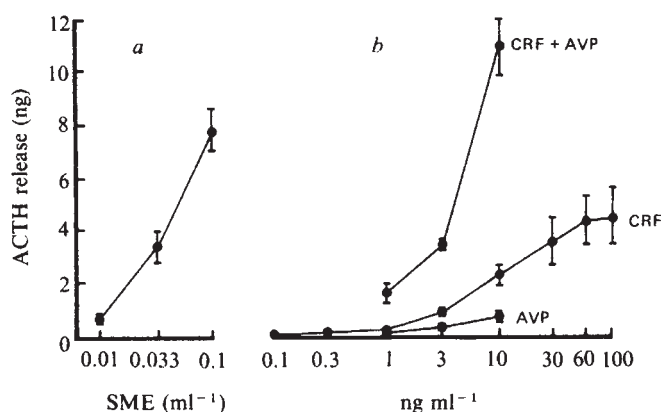


Fig. 1 Rat isolated pituitary cells, prepared by gentle mechanical agitation in Earle's balanced salt solution (EBBS; Gibco Biocult) containing trypsin (2.5 mg ml⁻¹), were mixed with swollen BioGel P2 (200–400) mesh, packed into a 2-ml plastic column and perfused (0.5 ml min⁻¹) with EBBS containing human serum albumin (0.25%), streptomycin (25 µg ml⁻¹), benzylpenicillin (15 µg ml⁻¹) Trasylol (100 kallikrein inactivator units per ml) and ascorbic acid (60 µg ml⁻¹). No significant losses of vasopressin or ACTH occur in passage of these peptides through the system. CRF bioactivity was measured by ACTH release in excess of background as monitored by immunoassay of 1-ml (2 min) fractions of cell column effluent. *a*, Rat SME (1 SME = 3–4 mg wet weight tissue¹²); *b*, synthetic vasopressin and 41 residue ovine hypothalamic CRF (Universal Biologicals Ltd, Cambridge) tested separately and simultaneously. Results are mean ± s.e.m. of *n* = 3 experiments.

exceeding the additive responses 4-fold for untreated SME compared with 2.4-fold for vasopressin antibody-quenched SME. This suggests that a considerable part of the CRF's activity was due to its interaction with vasopressin. The residual potentiation could be due to interaction with remaining free vasopressin or with a completely different factor. In our original antibody quenching experiments¹³, we did not include ascorbic acid in the media as its importance was not then appreciated. In more recent experiments performed up to 2 yr later, with ascorbic acid in the media, dilutions of antiserum which had been capable of quenching SME CRF activity below the sensitivity of our bioassay, now produced a 60–70% reduction of activity. It is impossible to deduce whether this is due to a deterioration of the antiserum so that all the vasopressin was not fully quenched, or, as we feel more probable, the ascorbic acid stabilizes a third factor^{9,14}. This point was clarified by using SME from the Brattleboro rat which is completely deficient in vasopressin and which has reduced ACTH-releasing activity^{13,15}. A significant potentiation was observed when synthetic CRF was added to Brattleboro SME (Fig. 3) although the extent of potentiation was again considerably less than that observed with untreated normal SME. Therefore it appears that the new CRF is synergizing with other factors apart from vasopressin in the CRF complex.

We have already presented some evidence that a molecule with free thiol groups is necessary for full CRF bioactivity of SME¹⁴ and that vasopressin in some way confers stability on the CRF complex¹³. It is unlikely that vasopressin interacts with the cysteine-free 41 residue hypothalamic peptide, so we postulate that the CRF complex consists of vasopressin, the 41 residue CRF plus another factor(s).

The multi-molecular nature of CRF has been reported by ourselves¹⁰ and by other workers^{7–9}, but the identity of one of the factors as vasopressin is controversial^{9,14}. Evidence in favour of the role of vasopressin includes immunohistochemical identification of vasopressinergic neurones terminating at the portal capillary bed of the zona externa of the median eminence, which respond to manipulations of the pituitary–adrenal axis^{16–18}; high levels of vasopressin in portal blood¹⁹; observations from this laboratory using the perfused isolated rat anterior pituitary cell column bioassay coupled with chromatography^{2,9,10,12–14}; and potentiation of Brattleboro rat CRF by addition of synthetic vasopressin^{13,20}. Furthermore, CRF activity from porcine and ovine hypothalamic extracts which chromatographs as vasopressin has the amino acid com-

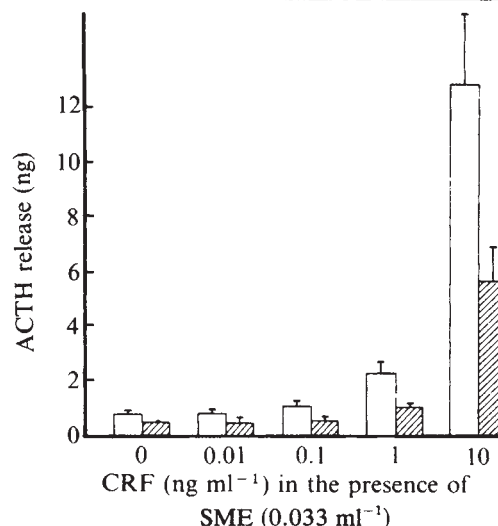


Fig. 3 CRF bioactivity of varying doses of synthetic CRF combined with normal (□) and Brattleboro (▨) rat SME (0.033 SME ml⁻¹) ($n = 4$).

position of lysine-vasopressin LVP⁹ and arginine-vasopressin¹¹ respectively. Despite some evidence to the contrary^{7,21} (see ref. 9 for review), a role for vasopressin as a physiological CRF is now being reconsidered more carefully^{15,22}.

The 41 residue peptide of ovine hypothalamic origin recently characterized by Vale *et al.*¹¹ as a potent CRF has enabled us to investigate its interaction with vasopressin. We found that although the new CRF is more potent in releasing ACTH in our bioassay than is vasopressin, it is still not as active a CRF at the rat anterior pituitary corticotrope as is rat SME. However, we have demonstrated a dramatic synergism between vasopressin and the new CRF over the linear portion of their individual log dose–response curves. Although the resultant curve is qualitatively identical to that for an extract of the rat stalk median eminence region, the amount of peptides required appear to be relatively high. In 0.1 SME, we would expect to find 1 ng immunoactive vasopressin, whereas recombination of 1 ng ml⁻¹ of both vasopressin and CRF produces a response far less than that observed for 0.1 SME ml⁻¹ (Fig. 1). Until a radioimmunoassay for the 41 residue CRF is established and the amounts of the new CRF (or related substance) in our extract can be measured, the concentration of the synthetic CRF used in our experiments remains arbitrary. On the basis of our chromatographic evidence that CRF consists of vasopressin plus at least two other synergizing factors, we previously added vasopressin to normal rat SME, but found only additive responses, suggesting that the other factors are present in limiting amounts. This was borne out to a large extent when vasopressin added to Brattleboro rat SME in amounts found in normal SME significantly potentiated CRF bioactivity¹³. Furthermore, potentiation of CRF bioactivity by addition of synthetic 41 residue CRF to normal SME supports this idea and suggests that a peptide similar to the synthetic CRF may play a significant part in controlling ACTH secretion in the rat. From the results with vasopressin antiserum-quenched SME and vasopressin-free SME from Brattleboro rats (Figs 2, 3) we deduce that at least one other substance besides vasopressin and the 41 residue CRF will be required for full CRF bioactivity in our system. We have evidence that, apart from vasopressin and a CRF which co-elutes with it on Sephadex G-50 (but not on Biogel P2), the rat and pig hypothalamic CRF complex contains a material of apparent molecular weight 5,000–8,000 which can be generated from a higher molecular weight precursor and which potentiates the CRF activity of vasopressin⁹. An antiserum to the 41 residue CRF may allow us to study its relationship to these CRFs and to the many other reported partially purified CRFs (see ref. 9 for review).

This work was supported by the MRC.

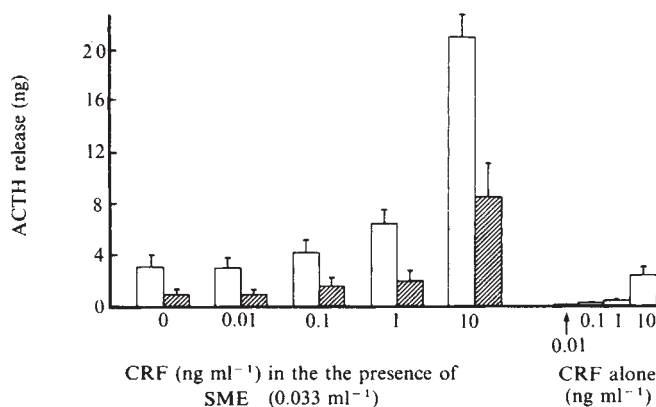


Fig. 2 CRF bioactivity of different doses of synthetic CRF combined with rat SME (0.033 SME ml⁻¹) compared with synthetic CRF alone. SME had been preincubated overnight at 4°C in bioassay perfusion buffer in the presence (▨) and absence (□) of vasopressin antiserum² (final dilution 1:1,000). Vasopressin antiserum had no effect on synthetic CRF given alone. As controls (not illustrated), normal rabbit serum and a somatostatin antiserum (final dilution 1:1,000) were substituted for vasopressin antiserum. Responses were superimposable on those for synthetic CRF and SME preincubated in the absence of vasopressin antiserum (□) with no effect on the potentiating ability of synthetic CRF. Results are mean $\pm$ s.e.m. ($n = 5$).

Received 11 January; accepted 28 July 1982.

1. Burgus, R. & Guillemin, R. A. *Rev. Biochem.* **39**, 499–526 (1970).
2. Gillies, G., Greidanus, T. B. V. W. & Lowry, P. J. *Endocrinology* **103**, 528–534 (1978).
3. Pearlmutter, A. F., Dokas, L., Kong, A., Miller, R. & Saffran, M. *Nature* **283**, 698 (1980).
4. Royce, P. C. & Sayers, G. *Proc. Soc. exp. Biol. Med.* **142**, 842–845 (1960).
5. Schally, A. V. & Guillemin, R. *Proc. Soc. exp. Biol. Med.* **112**, 1014 (1963).
6. Schally, A. V., Arimura, A. & Kastin, A. J. *Science* **179**, 341–350 (1973).
7. Pearlmutter, A. F., Rapino, E. & Saffran, M. *Endocrinology* **97**, 1336–1339 (1975).
8. Sayers, G., Hanzmann, E. & Bodansky, M. *FEBS Lett.* **116**, 236–238 (1980).
9. Gillies, G. & Lowry, P. J. in *Frontiers in Neuroendocrinology* Vol. 7 (eds Martini, L. & Ganong, W. F.) 45–74 (Raven, New York, 1982).
10. Gillies, G. & Lowry, P. J. *Nature* **278**, 463–464 (1979).
11. Vale, W., Spiess, J., Rivier, C. & Rivier, J. *Science* **213**, 1394–1397 (1981).

12. Gillies, G. & Lowry, P. J. *Endocrinology* **103**, 521–527 (1978).
13. Gillies, G. & Lowry, P. J. *J. Endocr.* **84**, 65–73 (1980).
14. Gillies, G. & Lowry, P. J. in *Interactions within the Brain-Pituitary-Adrenocortical System* (eds Jones, Gilham, Dallman & Chattopadhyay) 51–61 (Academic, New York, 1979).
15. Sokol, H. W., & Zimmerman, E. A. *Ann. N.Y. Acad. Sci.* (in the press).
16. Watkins, W. B., Schwabedal, P. & Bock, R. *Cell Tissue Res.* **152**, 411–421 (1974).
17. Zimmerman, E. A. in *Frontiers of Neuroendocrinology* Vol. 4 (eds Martini, L. & Ganong, W. F.) 25–62 (Raven, New York, 1976).
18. Zimmerman, E. A. *et al. Ann. N.Y. Acad. Sci.* **297**, 405–419 (1977).
19. Goldman, H. & Lindner, L. *Separatum Experientia* **18**, 279–281 (1962).
20. Buckingham, J. C. *J. Physiol., Lond.* **312**, 9–16 (1981).
21. Lutz-Butcher, B., Koch, B., Miahle, C. & Briaud, B. *Neuroendocrinology* **30**, 178–182 (1980).
22. Baertisch, A. in *Neuroendocrinology of Vasopressin, Corticotropin and Opiomelanocortin* (eds Baertisch, A. & Driefuss, J.-J.) (Academic, London, in the press).

Oligodendrocyte abnormalities in *shiverer* mouse mutant are determined in primary chimaeras

K. Mikoshiba*, M. Yokoyama†, Y. Inoue‡, K. Takamatsu*, Y. Tsukada* & T. Nomura†

* Department of Physiology, Keio University School of Medicine, 35 Shinano-machi, Shinjuku-ku, Tokyo 160 Japan

† Central Institute for Experimental Animals, 1430 Nogawa, Miyamae-ku, Kawasaki 213, Japan

‡ Department of Anatomy, Hokkaido University, School of Medicine, 15–7, Kita-ku, Sapporo, Hokkaido, Japan

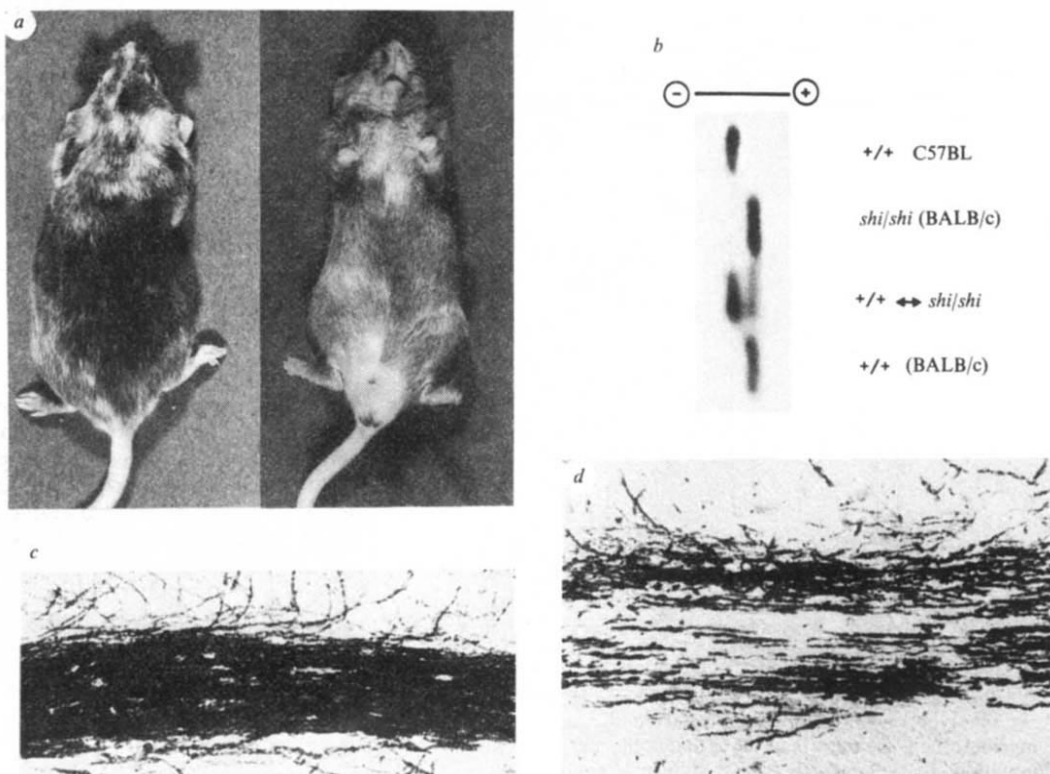
There is increasing interest in hereditary dysmyelinating mutant mice such as *jimpy*, *quaking*, *shiverer* and *twitcher*, as animal models for leukodystrophy in man^{1–3}. These mutants are characterized by severe tremor and tonic convulsions. *Shiverer* is an autosomal recessive mutant which displays abnormal and poor myelination of the central nervous system^{4–9} and is genetically different from the other dysmyelinating mutants. Despite many reports on the morphological and chemical changes which occur in *shiverer* mice, the pathogenesis of the abnormal and poor myelin formation is unknown. We have now produced chimaeras by aggregation of eight-cell-stage embryos from wild-type and *shiverer* mutants. Analysis of the white matter of the chimaera suggests that the oligodendrocyte is responsible for the abnormal and poor myelination in *shiverer*.

Chimaeras were produced by the aggregation method^{10–13}, which makes it possible to know the interactions among cells. To facilitate the judgement of mosaicism of different genotypes in the brain, C57BL/6N mouse strain (black coat colour) was used as wild-type control, as they have a different isozyme pattern of glucose phosphate isomerase (GPI) which is distinct from that of *shiverer* mutant mice on the BALB/c background (white coat colour). We have obtained six chimaeras, and selected for analysis the chimaera shown in Fig. 1a. Examination of the strain-specific GPI isozyme pattern of brain tissues in the chimaera shows chimaerism of both genotypes as well as coat colour. The gel electrophoretic pattern of GPI isozyme revealed that brain tissue in the chimaera was composed of isozymes of both BALB/c and C57BL/6N (Fig. 1b). In the several *shiverer* mice examined so far, we have observed a clear correlation between the mosaicism of coat colour and that of the GPI isozyme pattern.

A neurochemical characteristic of the *shiverer* mutation is the absence of myelin basic protein (MBP), which is a major component of the myelin sheath. The immunohistochemical reaction using antiserum against MBP also shows the absence of MBP^{4,14}. Here we have used the immunohistochemical reaction of MBP as a marker to distinguish *shiverer*-type myelin from that of the wild-type control.

If all myelin formation was abnormal in the chimaera central nervous system, we would suspect the presence of an extrinsic harmful factor derived from the *shiverer* which could not be neutralized by the normal tissue. Conversely, if myelin formation in the central nervous system of the chimaera was similar

Fig. 1 Mosaics of coat colour, brain tissue GPI isozyme pattern and brain MBP-positive and -negative fibres from the normal-*shiverer* chimaera mouse. *a*, Chimaera mouse (*shi/shi* ↔ *+/+*) produced by the aggregation technique. *b*, Cellulose acetate electrophoresis of mouse strain-specific enzyme variants of GPI from homogenates of part of the brain. The chimaera was produced from the *+/+* (C57BL) and *shi/shi* (BALB/c) mice. *c*, Immunohistochemical staining of the white matter of the control (*+/+*) cerebellum using anti-MBP serum. *d*, Immunohistochemical staining of the white matter of the chimaera by anti-MBP serum. MBP-positive fibres and MBP-negative fibres are mixed in the white matter of the chimaera, indicating the absence of humoral factors which might cause abnormal and poor myelin formation.



to that of the control, we would assume that the abnormal myelin formation of the *shiverer* had been resolved by the diffusion of some extrinsic humoral factor from the normal tissue to the *shiverer*, or by the removal by the control tissue of an extrinsic humoral factor responsible for abnormal myelin formation in *shiverer*. If the mutation were expressed intrinsic to the *shiverer* cells themselves, it seems probable that both normal and abnormal *shiverer*-type myelin would occur as a mixture or as patches.

In fact, we observed MBP-positive and -negative fibres mixed in the white matter (Fig. 1d, compare with Fig. 1c, which is a control section), thus eliminating the possibility that humoral factors caused the abnormal myelin formation of the *shiverer*. We therefore suggest that the *shiverer* mutation occurs intrinsic to the cells themselves, namely neuronal or glial cells.

Myelination is a process of interaction of neuronal and glial cells. Oligodendrocytes send out their processes to search for the axon to myelinate³. There are thus two possible causes of dysmyelination in *shiverer* mice: abnormalities in the neuronal cells which are to be myelinated or abnormalities in the glial cells responsible for the myelination. In fact, morphological abnormalities have been detected in both the neurone, using the silver impregnation technique¹⁵, and the glial cell (abnormal myelin formation in oligodendrocytes^{5,8,9,16} and a hypertrophy of astrocytes¹⁴).

If the cause of dysmyelination is primarily on the neuronal side, *shiverer*-type axon would always have *shiverer*-type myelin (Fig. 2b) and normal axon would have normal myelin (Fig. 2a). In such a case, the type of myelin observed in Fig. 2c would not be present. However, if dysmyelination in *shiverer* is due to an abnormality of the oligodendrocyte rather than the neurone, *shiverer*-type myelin would be seen adjacent to normal myelin on the axon of either control or *shiverer* origin. The type of myelin formation observed in Fig. 2c would then be present as well as that seen in Fig. 1a and b. Figure 3 shows abnormal *shiverer*-type myelin, with the major dense line absent, lying adjacent to normal myelin on the same axon; this feature is often observed under the electron microscope.

We therefore suggest that the cause of abnormal myelination in *shiverer* mutant mouse is not of neurogenic origin, but rather that the oligodendrocyte itself is responsible.

Chimaeric analysis has proved successful in analysing the pathogenesis of the photoreceptor cells degeneration in the retinal dystrophy mutant¹⁷, the degeneration of Purkinje cells in the Purkinje cell degeneration mutant¹⁸, in muscular dystrophy¹⁹ and also in other neurological mutants²⁰. The

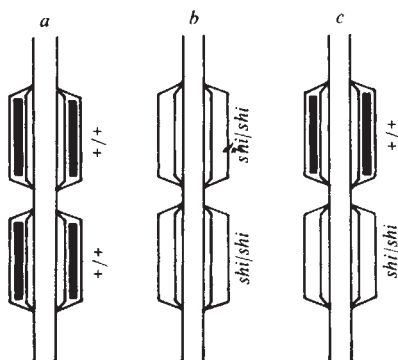


Fig. 2 Three hypothetical models of myelin formation in the central nervous system of the *shiverer*-normal chimaera. The dense bar in the myelin indicates the major dense line which is absent in the *shiverer*-type myelin. If the *shiverer* mutation acts on the axons in the chimaera, all the myelin formed on the axon originated from *shiverer* would be *shiverer*-type (b), and the myelin sheathes on the control axon would all be normal (a). If the *shiverer*-type myelin is observed adjacent to the control-type myelin (c), it is the oligodendrocyte that is affected by the *shiverer* mutation. In the *shiverer*-normal chimaera, all three types of myelin formation were observed.

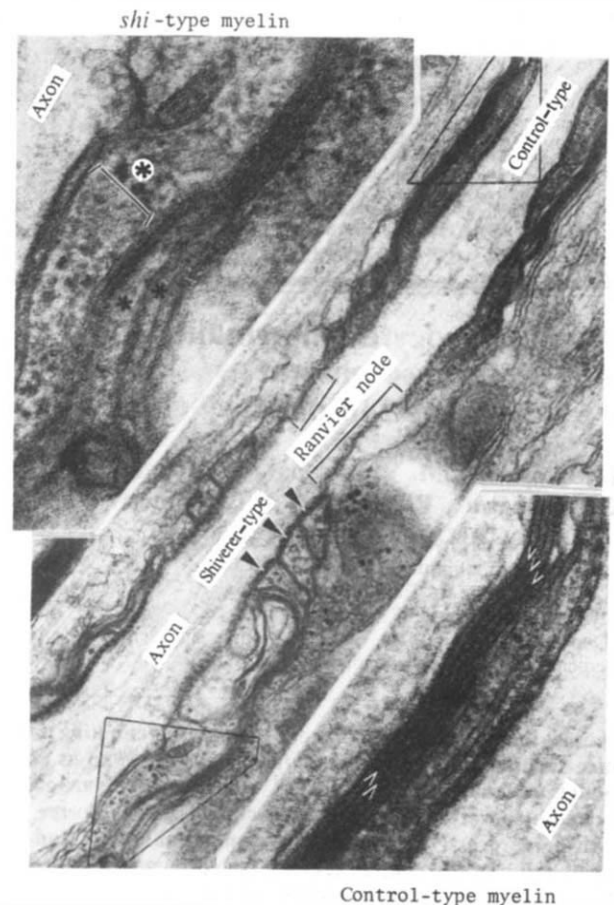


Fig. 3 An axon with control and *shiverer*-type myelin in the white matter of the chimaera mouse. Electron microscopic demonstration of the sagittal section of the myelinated axon in the corpus callosum. In the typical myelin lamella, major dense lines (indicated by small arrowheads) are clearly observed (lower inset) and paranodal cytoplasmic pockets lose their cytoplasm immediately to form the major dense line (central figure), while in the myelin of the lower left (central figure) and upper inset, which is believed to be of *shiverer* origin, the paranodal cytoplasmic pockets are enlarged and irregular in form and isolated pockets (indicated by larger arrow heads) are often observed. In the *shiverer*-type myelin, cytoplasm of the oligodendrocyte (indicated by *) is found where the major dense line would form in normal myelin.

present results demonstrate the site at which the *shiverer* mutation is acting, and future work should establish whether the inability of the oligodendrocyte to synthesize MBP is controlled by other cells (for example, astrocytes) through short-distance diffusible factors or by fibres attached to oligodendrocytes in *shiverer*. Double immunohistochemical staining using antisera against MBP and strain-specific GPI should elucidate whether or not MBP-synthesizing activity in control oligodendrocytes (the origin of which can be detected by the GPI antiserum) is converted to *shiverer*-type (which can be detected by MBP antiserum) by the adjacent *shiverer* cells.

We thank Drs F. Lachapelle and N. Baumann for their generous supply of *shiverer* mutant mice, and Dr J.-L. Guenet, Dr M. Katsuki and N. Ohsawa for valuable discussions.

Received 20 April; accepted 21 July 1982.

1. Baumann, N. *Neurological Mutations Affecting Myelination, Research Tools in Neurobiology Correlations to Human Neurological Diseases* (Elsevier, Amsterdam, 1980).
2. Sidman, R. L., Green, M. C. & Appel, S. H. *Catalog of the Neurological Mutants of the Mouse* (Harvard University Press, Cambridge, 1965).
3. Morell, P. *Myelin* (Plenum, New York, 1977).
4. Dupouey, O. et al. *Neurosci. Lett.* **12**, 113-118 (1979).
5. Privat, A., Jacque, C., Bourre, J. M., Dupouey, P. & Baumann, N. *Neurosci. Lett.* **12**, 107-112 (1979).
6. Mikoshiba, K., Nagaike, K., Kohsaka, S., Takamatsu, K. & Tsukada, Y. *Bull. Japan. neurochem. Soc.* **18**, 200-203 (1979).

7. Mikoshiba, K., Aoki, E. & Tsukada, Y. *Brain Res.* **192**, 195–204 (1980).
8. Mikoshiba, K., Nagaike, K. & Tsukada, Y. *J. Neurochem.* **35**, 465–470 (1980).
9. Inoue, Y., Nakamura, R., Mikoshiba, K. & Tsukada, Y. *Brain Res.* **219**, 85–94 (1981).
10. Mintz, B. *Science* **138**, 594–595 (1962).
11. Mintz, B. *Am. Zool.* **2**, 432 (1962).
12. Mintz, B. in *Methods in Mammalian Embryology* (ed. Daniel, J. C. Jr.) 186–214 (Freeman, Toppan 1971).
13. Mullen, R. J. & Whitten, W. K. *J. exp. Zool.* **178**, 165–176 (1971).
14. Mikoshiba, K., Takamatsu, K., Kohsaka, S., Tsukada, Y. & Inoue, Y. in *Genetic Approaches to Developmental Neurobiology* (ed. Tsukada, Y.) 195–221 (Tokyo University Press, 1982).
15. Friedrich, V. L., Koniecki, D. L. & Massa, P. T. in *Neurological Mutations Affecting Myelination* (ed. Baumann, N.) 141–146 (Elsevier, Amsterdam, 1980).
16. Rosenbluth, J. J. *comp. Neurol.* **194**, 639–648 (1980).
17. Mullen, R. J. & LaVail, M. M. *Science* **192**, 801–799 (1976).
18. Mullen, R. J. in *Approaches to the Cell Biology of Neurons Vol. 2* (eds Cowan, W. M. & Ferrendelli, J. A.) 47–65 (Society for Neuroscience, Bethesda, Maryland, 1976).
19. Peterson, A. C. *Nature* **248**, 561–564 (1974).
20. Bray, G. M., Rasminsky, M. & Aguayo, A. J. *Rev. Neurosci.* **4**, 127–162 (1981); *Neurology* **31**, 118 (1981); *Soc. Neurosci. Abstr.* **7**, 545 (1981).

A Cl^- conductance activated by hyperpolarization in *Aplysia* neurones

Dominique Chenoy-Marchais

Laboratoire de Neurobiologie, École Normale Supérieure,
46 rue d'Ulm, 75005 Paris, France

Although many voltage-gated cation channels have been described and extensively studied in biological membranes, there are very few examples of voltage-gated anion channels. Chloride conductances activated by depolarization have been observed in skate electroplaque^{1,2} and in frog and chick skeletal muscle^{3–5}. A Cl^- conductance activated by hyperpolarization has been suggested both for frog muscle treated with acid (pH 5) solutions³, and for crayfish muscle^{6,7} where it could account for the fact that the pronounced inward-going rectification of the I - V curve disappears if the fibres have been soaked in a Cl^- -free solution. More recently, voltage-dependent anion channels extracted from biological membranes have been incorporated^{8–10} into artificial membranes. I now report that in *Aplysia* neurones, and in particular those in which the internal Cl^- concentration has been increased, a Cl^- conductance can be observed which is slowly activated by hyperpolarization and shows a very steep voltage dependence. This time- and voltage-dependent Cl^- conductance probably exists also in many other cells. Its presence might explain why it is difficult when using KCl-filled microelectrodes to maintain prolonged hyperpolarizations¹¹. This Cl^- conductance constitutes a new type of inward-going rectification distinct both from the classical 'anomalous rectification' which involves selective K^+ channels^{12,13} and from the current termed i_i in heart muscle that is presently attributed to a cationic conductance¹⁴.

Most experiments were performed on identified *Aplysia* neurones (the A neurones of the cerebral ganglion) which were voltage-clamped using two intracellular microelectrodes filled with either K_2SO_4 (0.5 M) or KCl (1–3 M)¹⁵. These neurones respond to acetylcholine by a rapid conductance increase specific for Cl^- ions and a slower conductance increase specific for K^+ ions¹⁵: the reversal potentials of these two responses provide direct measurements of the Cl^- and K^+ equilibrium potentials (E_{Cl} and E_{K}). The neurones were held near the resting potential, at -45 or -50 mV, and were then hyperpolarized for a few seconds by a square voltage jump to -100 mV. When the intracellular electrodes were filled with K_2SO_4 , the current change during the pulse was usually square, although sometimes the inward current recorded at -100 mV showed a slight and slow increase during the pulse (Fig. 1a). On the other hand, when the recording electrode was a leaky electrode filled with KCl, or when chloride ions were injected iontophoretically through an additional intracellular KCl electrode, I observed, after the hyperpolarizing jump, in addition to the instantaneous current change, a very large inward current which increased slowly for a few seconds. A large slow tail current was observed after repolarization from -100 to -50 mV. The fact that at -100 mV all the currents crossing the ionic channels of the

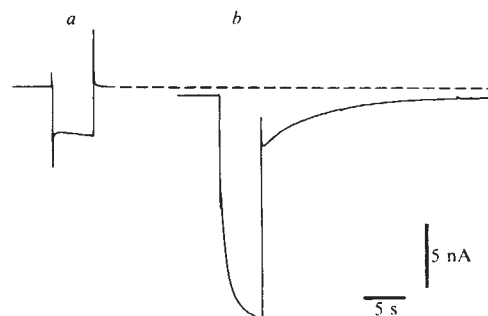


Fig. 1 Increasing the intracellular Cl^- concentration reveals a large slow hyperpolarization-induced inward current which slowly deactivates on repolarization. The cell was impaled with two microelectrodes filled with K_2SO_4 (0.5 M) plus one double-barrelled micropipette containing K_2SO_4 (0.5 M) in one barrel and KCl (1 M) in the other. The cell was maintained at -50 mV between hyperpolarizing pulses (5 s to -100 mV) which were applied every 3 min. The current recorded during such a voltage jump is shown before (a) and after (b) the internal Cl^- concentration had been raised. In a, a braking current of 10 nA was applied to the double-barrelled micropipette to avoid leakage of Cl^- into the cell; in b, Cl^- ions had been injected iontophoretically with a current of 30 nA for 15 min. The steady-state current at the holding potential was slightly altered by the chloride injection, suggesting that, after the injection, the voltage-dependent inward current in this cell was slightly activated at -50 mV.

membrane were inward currents (all the ionic equilibrium potentials were more depolarized than -100 mV), suggested that the slow inward current observed at -100 mV was due to the slow activation of an ionic conductance induced by hyperpolarization and that the tail current was due to the slow deactivation of this conductance.

The ionic specificity of these slow voltage-dependent currents was investigated. As larger currents could be obtained by loading the neurones with Cl^- , I used systematically a KCl voltage recording microelectrode and only started the analysis after E_{Cl} had reached a stable value. The currents activated by the hyperpolarizing jumps were clearly not due to the activation of a selective K^+ conductance, as in cells in which E_{K} was measured precisely (as the reversal potential of the slow cholinergic response), large slow increases in inward current were triggered by voltage jumps from -50 mV to E_{K} , and slow tail currents were observed after repolarization from -100 mV to E_{K} (E_{K} was always found very close to -80 mV). I considered the possibility that the current is similar to the i_i of heart cells (ref. 14; see also ref. 16): these currents, which resemble the current described here in that they are activated by hyperpolarization, are attributed to an increased cationic conductance to both Na^+ and K^+ and have been shown to be blocked by Cs^+ . Neither of these properties applies to the current observed in *Aplysia* neurones: (1) Complete substitution of external Na^+ by Tris, which is usually not permeant through cationic channels, did not induce any change in the current, either in amplitude or in kinetics. (2) Addition of Cs^+ ions (1–50 mM) did not modify the current (provided Co^{2+} ions (10 mM) were present both in control and Cs^+ -containing solutions to prevent Cs^+ -induced changes in synaptic activity). (3) In two experiments, slow hyperpolarization-activated inward currents very similar to those usually observed in control seawater, were recorded in an isotonic CsCl extracellular solution, after replacement (by treatment with nystatin¹⁷) of intracellular monovalent cations by Cs^+ ions. (The extracellular CsCl solution was Na^+ -free and K^+ -free (517 mM CsCl , 10 mM CaCl_2 , 50 mM MgCl_2 , 5 mM HEPES- CsOH , pH 7.6). For treatment with nystatin, the ganglion was perfused with a Cs^+ -loading solution (300 mM CsCl , 394 mM sucrose, 100 mM MgSO_4 , 5 mM HEPES- CsOH); nystatin (20 mg l^{-1}) was applied in this solution for 30 min. The ganglion was then washed for 30 min with the nystatin-free Cs^+ -loading solution before being perfused with the isotonic CsCl solution; before recording, the cell bodies were isolated

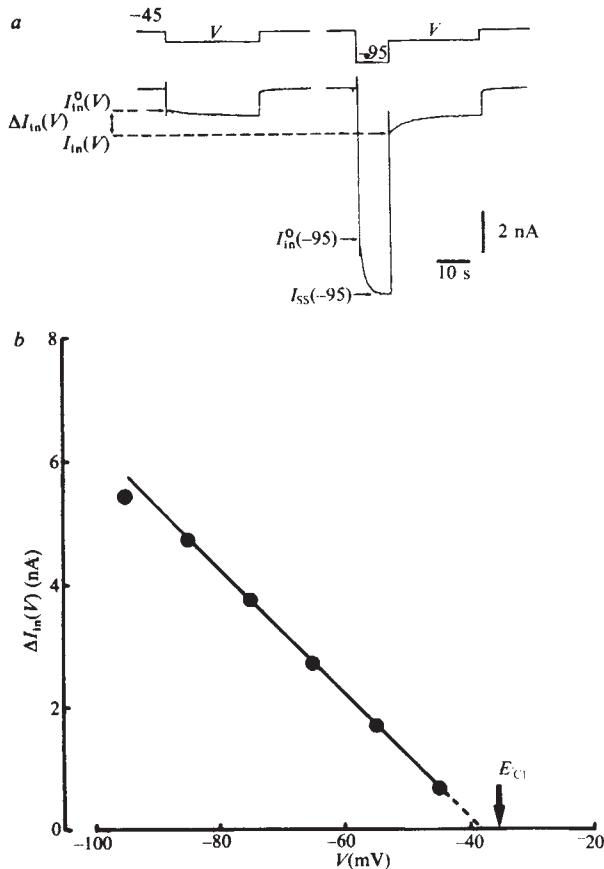


Fig. 2 Extrapolation of the instantaneous current-voltage curve to zero current gives a reversal potential close to E_{Cl} (b). In a are shown the voltage jumps which were used to establish the curve in b (the records presented come from an experiment similar to that reported for b). Arrows in a indicate the measured current values. $I_{in}^0(V)$ is the instantaneous current at V after a hyperpolarizing jump from -45 mV to V. At the holding potential (-45 mV) the current under investigation was not activated: pulses from -35 to -45 mV did not induce any slow time-dependent inward current at -45 mV. Thus, $I_{in}^0(V)$ is the 'resting' current at V. $I_{in}(V)$ is the instantaneous current at V after a depolarizing jump from a hyperpolarized level (-95 mV) at which the inward current was activated for 10 s. ($I_{in}^0(V)$ and $I_{in}(V)$ were measured by extrapolation to time zero of the exponential kinetics of the currents recorded at V.) $\Delta I_{in}(V)$, the difference between $I_{in}(V)$ and $I_{in}^0(V)$, is plotted in b. E_{Cl} (indicated by an arrow) was measured as the reversal potential of the Cl^- response to ionophoretically applied carbachol. It was not modified by the hyperpolarizing pulses applied in this experiment. $I_{ss}(V)$ is the steady-state value of the current at V.

by cutting the axons in order to avoid synaptically-induced changes in conductance. The intracellular microelectrodes were filled with CsCl (0.5 M) or CsSO₄ (0.5 M).

These results, together with the fact that the currents were very sensitive to the intracellular Cl^- concentration (Fig. 1) suggested that the observed currents were due to a voltage-gated Cl^- conductance. This was further demonstrated by comparing the reversal potential of the fast cholinergic response (E_{Cl}) with the reversal potential (E_r) of the tail currents observed after repolarization to various membrane potentials after a given hyperpolarizing pulse. Direct measurement of this reversal potential was difficult because in order to have large hyperpolarization-induced currents, it was necessary to load the cell with Cl^- thus bringing E_{Cl} within a range (more depolarized than -40 mV) where many currents are activated by a sudden depolarization. (The problem was further compounded by the fact that the kinetics of the tail currents becomes rapid in this range (see Fig. 3a).) The reversal potential of the

tail currents was therefore estimated by extrapolating values measured at various levels below the reversal potential (Fig. 2). The extrapolated reversal potential (E_r) was very close to E_{Cl} ($E_r - E_{Cl} = -2.3 \pm 2.0$ (7) (mean $\pm$ s.d. (n)). Furthermore, in four experiments in which half the extracellular chloride was replaced by isethionate, both E_r and E_{Cl} shifted by similar values, that is, 10 ± 2 mV and 9.3 ± 0.9 mV, respectively. In another experiment, the intracellular Cl^- concentration was increased in two steps, by two successive iontophoretic injections of Cl^- ; E_{Cl} shifted from -30 mV to -20 mV while E_r shifted from -32 mV to -22 mV.

Thus, the hyperpolarization-induced slow inward current and the corresponding tail current appear to result from a voltage-gated Cl^- conductance slowly activated by hyperpolarization and slowly deactivated by repolarization.

To characterize the activation and deactivation of this Cl^- conductance, I analysed the kinetics and the amplitude of the ON relaxations induced by square hyperpolarizing pulses from -50 mV to various test potentials, and the kinetics of the OFF relaxations induced by repolarization to a variable test potential after a 10-s hyperpolarizing pulse from -50 to -100 mV. In some experiments, these relaxations could be described by single exponentials. (In other experiments which were disregarded, the relaxations presented an additional slower component which may have resulted from an imperfect clamp of the membrane potential). The time constants of the exponential relaxations ($\tau_{ON}(V)$ and $\tau_{OFF}(V)$) are plotted in Fig. 3a. For membrane potentials which allowed accurate measurement of both $\tau_{ON}(V)$ and $\tau_{OFF}(V)$ the values of these two time constants were found to be identical. This was confirmed in six other such experiments. Figure 3a shows that the kinetics of activation is much more rapid at -100 mV than at -70 mV, varying by a factor e per 14 mV. Figure 3b illustrates the amplitude of the conductance change during each hyperpolarizing pulse. In this cell, the voltage-gated Cl^- current did not seem to be activated at -50 mV as a pulse from -40 to -50 mV did not trigger any slow time-dependent inward current. (Preliminary experiments (not shown) indicate that the threshold for activation of this Cl^- conductance becomes lower as the intracellular Cl^- concentration is increased. This may explain why in the experiment of Fig. 1, the Cl^- conductance is already activated at -50 mV, whereas this is not so in the experiment of Fig. 3.) Figure 3b therefore gives the voltage-dependence of the steady-state Cl^- conductance. It is very steep between -50 and -70 mV (e -fold increase per 9 mV) and saturates at highly hyperpolarized membrane potentials.

Most of the results presented here were obtained on intact A cells, but similar results were obtained from cell bodies isolated by cutting off the axons. The Cl^- conductance described

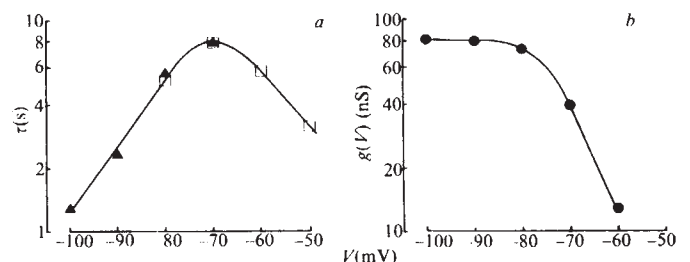


Fig. 3 Voltage-dependence of the kinetics of activation and deactivation and of the steady-state conductance. a, $\tau_{ON}(V)$ ($\blacktriangle$) and $\tau_{OFF}(V)$ ($\square$); b, steady-state conductance $g(V)$. $g(V)$ was calculated as follows:

$$g(V) = \frac{I_{ss}(V) - I_{in}^0(V)}{V - E_{Cl}}$$

where $I_{ss}(V)$ is the steady-state value of the current at the potential V and $I_{in}^0(V)$ is the instantaneous current at V after a jump from a potential at which the Cl^- conductance was not activated. In the experiment illustrated, the jump started from -50 mV; E_{Cl} was -35 mV.

is therefore not synaptic and is at least partially located on the soma. Preliminary results indicate that this voltage-gated Cl^- conductance described in A cells exists in many other *Aplysia* neurones.

In conclusion, I have described a neuronal non-synaptic Cl^- conductance in which both the steady-state conductance and the kinetics of activation are highly voltage-dependent (Fig. 3) while the instantaneous current-voltage curve is approximately linear (Fig. 2). These properties resemble those of some voltage-gated Cl^- conductances previously described in non-neuronal systems^{3,7,10}. It is possible that such a conductance exists in many other neurones. It may coexist with other inward rectifying conductances like the classical 'anomalous rectification' of the K^+ -conductance. Indeed, two different mechanisms have already been suggested¹⁸ to account for the anomalous rectification of molluscan neurones¹⁹. The functional role of this Cl^- conductance may be partly to stabilize membrane potential and prevent excessive hyperpolarization. It could also be a mechanism of regulating the intracellular Cl^- concentration, helping to maintain it at low levels in cells which, like *Aplysia* neurones, are normally very impermeant to Cl^- (ref. 20) but which may become loaded with Cl^- during repeated activation of Cl^- inhibitory synapses.

I thank Professor P. Ascher for his encouragement and help during this work and D. Paupardin-Tritsch for isolation of cell bodies. This work was supported by grants from the CNRS, the DGRST and the Université Pierre et Marie Curie.

Received 13 April; accepted 7 July 1982.

- Grundfest, H., Aljure, E. & Janiszewski, L. *J. gen. Physiol.* **45**, 598 A (1962).
- Hille, B., Bennett, M. V. L. & Grundfest, H. *Biol. Bull.* **129**, 407-408 (1965).
- Warner, A. E. *J. Physiol., Lond.* **227**, 291-312 (1972).
- Fukuda, J. *Science* **185**, 76-78 (1974).
- Fukuda, J. *Nature* **257**, 408-410 (1975).
- Reuben, J. P., Girardier, L. & Grundfest, H. *Biol. Bull.* **123**, 509-510 (1962).
- Ozeki, M., Freeman, A. R. & Grundfest, H. *J. gen. Physiol.* **49**, 1335-1349 (1966).
- Schein, S. J., Colombini, M. & Finkelstein, A. J. *Membrane Biol.* **30**, 99-120 (1976).
- Colombini, M. *Nature* **279**, 643-645 (1979).
- White, M. M. & Miller, C. J. *biol. Chem.* **254**, 10161-10166 (1979).
- Coombs, J. S., Eccles, J. C. & Fatt, P. J. *Physiol., Lond.* **130**, 326-373 (1955).
- Adrian, R. H. *Prog. Biophys. molec. Biol.* **19**, 341-369 (1969).
- Ohmori, H., Yoshida, S. & Hagiwara, S. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4960-4964 (1981).
- Di Francesco, D. J. *Physiol., Lond.* **314**, 377-393 (1981).
- Ascher, P. & Chesnoy-Marchais, D. *J. Physiol., Lond.* **324**, 67-92 (1982).
- Adams, P. R. & Halliwell, J. V. J. *Physiol., Lond.* **324**, 62 P (1982).
- Russell, J. M., Eaton, D. C. & Brodwick, M. S. *J. Membrane Biol.* **37**, 137-156 (1977).
- Marmor, M. F. *J. Physiol., Lond.* **218**, 573-598 (1971).
- Tauc, L. & Kandel, E. R. *Nature* **202**, 1339-1341 (1964).
- Ascher, P., Kunze, D. & Neild, T. O. *J. Physiol., Lond.* **256**, 441-464 (1976).

Transplacental transfer of rodent microfilariae induces antigen-specific tolerance in rats

Azizul Haque & André Capron

Centre d'Immunologie et de Biologie Parasitaire, Institut Pasteur, 15, rue Camille Guérin, 59019-Lille, France

Microfilariae are the smallest form in the life-cycle of filarial nematode parasites. They are released by the adult female worms and migrate through the blood and extracellular fluids where they can be transmitted by vectors. A few reports have indicated the possibility of the transmission of microfilarial infection from mother to offspring¹⁻³. We have infected rats with adult females of the rodent filaria, *Dipetalonema viteae*, and report here that the transfer of *D. viteae* microfilariae does indeed occur through the placenta. Exposure to specific antigens early in development can readily induce immune tolerance⁴⁻⁶. We observed that a state of reversible immune unresponsiveness occurred in rats as a result of pre- and post-natal exposure to microfilariae and this was associated with impairment of T-cell responses. The induction of tolerance allowed *D. viteae* infective larvae to reach maturity in the Fischer rat which is otherwise innately resistant to this parasite.

All experiments were performed in inbred Fischer/ICO rats. Infant rats were bred in our animal facility and the adults were obtained commercially (IFFA CREDO). Infective larvae of *D. viteae* cannot reach maturity in Fischer rats but surgically transplanted female adults of *D. viteae* will establish and release microfilariae which circulate in the peripheral blood for about 120 days⁷.

In the present investigation, 10 mature female worms were collected under sterile conditions from hamsters infected 90 days previously and were transferred individually with blunt forceps or a probe within a few minutes of collection, into a subcutaneous pocket in the back of the anaesthetized female animals. A microfilaraemia was first detected on day 2, reaching a peak around day 20 to 30 and then declining and remaining at a low level until about days 120-130 after implantation. Seven days after infection, pairs of non-infected males and microfilaraemic females were mated and litters were born between 21 and 35 days later. All offspring irrespective of sex showed the presence of microfilariae in their peripheral blood. The number of microfilariae remained steady in the circulation until day 100 after birth and declined thereafter, disappearing from the peripheral blood at about days 120-130 after birth. The duration of microfilaraemia was much the same in the mothers and their offspring. The microfilaraemia resulting in the mother rats and in their offspring is shown in Table 1.

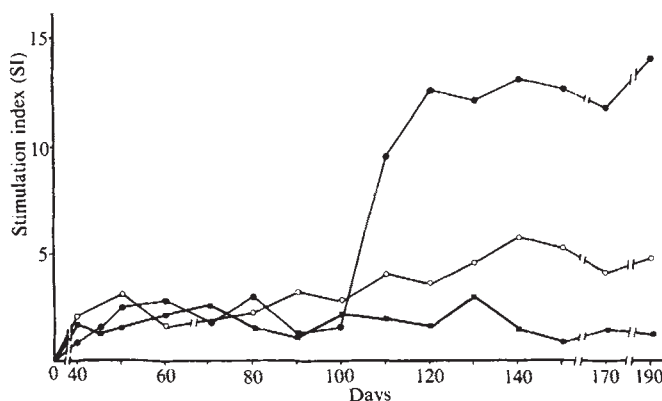


Fig. 1 *In vitro* lymphocyte proliferation in response to *D. viteae* antigens (prepared as described previously¹²) by spleen cells obtained from rats born to microfilaraemic (MR) or noninfected rats (NR). The degree of lymphocyte stimulation is expressed by the stimulation index (SI)

$$SI = \frac{\text{average c.p.m. in stimulated cultures}}{\text{average c.p.m. in unstimulated (control) cultures}}$$

The level of incorporation in unstimulated cultures was variable but within the range of $1,125 \pm 377$ to $3,322 \pm 764$ c.p.m. for normal spleen cells and within the range of $1,265 \pm 585$ to $3,698 \pm 798$ c.p.m. for spleen cells from infected animals. Each point represents the mean value and standard deviations were below 10% in all groups. Two rats were tested at each point. The optimal conditions for lymphocyte transformation towards parasite antigens, T and B cell mitogens were predetermined in both infected and noninfected Fischer rats (data not shown). Spleen cells were incubated for 96 h with parasite antigens ($250 \mu\text{g ml}^{-1}$). Proliferation response of spleen cells taken from MR either towards filarial antigens (●) or to *Ascaris lumbricoides* antigens (○). The response of spleen cells from NR toward filarial antigens (■) was measured at the same time. Glass bead fractionation was performed on each occasion. A few fibres of nylon-wool was placed into the barrel of a 5 ml glass syringe and cleaned glass beads ($150 \mu\text{m}$ size unispheres, Sigma) were added up to the 3 ml mark. It was then preincubated in RPMI plus 5% fetal calf serum (FCS) for 45 min at 37°C . Spleen cells (0.75×10^6) in 2 ml of the same medium, prewarmed to 37°C , were run into the column, which was incubated for a further 30 min at 37°C . Non-adherent cells were then eluted with 25 ml of warmed medium and were washed three times with RPMI only to remove FCS and finally resuspended in RPMI supplemented with the synthetic tripeptide glycyl-histidyl-lysine (20 ng ml^{-1}), L-glutamine ($29 \text{ mg per } 100 \text{ ml}$) and antibiotics. Lymphocyte transformation studies were performed in microtitre plates (Nunc) utilizing triplicate 0.1 ml cultures containing 0.5×10^6 spleen cells per culture well. Four hours before the end of the culture period $1 \mu\text{Ci}$ of [^3H]thymidine (1 Ci mmol^{-1} , CEA) was added to each culture. Microcultures were collected on a Titertek cell collector (Skatron). Filter disks were dried and transferred to scintillation vials with 10 ml of scintillation fluid (Lipo Luma) and assayed with a Nuclear Chicago liquid scintillation counter.

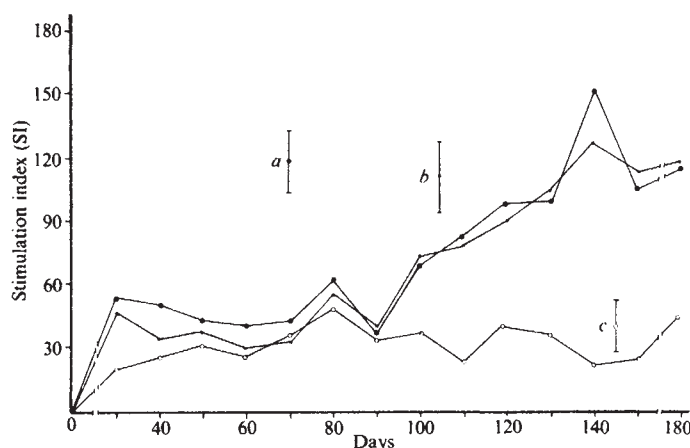


Fig. 2 The proliferative responses to PHA (*), Con A(●) and LPS(○) of spleen cells from rats born to microfilaraemic mothers, plotted as a function of time after infection. Spleen cells non-adherent to glass beads were used throughout as described in the legend to Fig. 1. SI values were calculated as described for Fig. 1. The level of incorporation in control cultures was variable but within the range of $1,052 \pm 309$ c.p.m. to $4,398 \pm 844$ c.p.m. for normal spleen cells and within the range of $1,013 \pm 586$ to $6,741 \pm 756$ for cells from infected animals. Each point represents the mean value from the rats and standard deviations were below 12% in all groups. Spleen cells (0.5×10^6) were incubated for 72 h with PHA, Con A ($100 \mu\text{g ml}^{-1}$) and for 96 h with LPS ($250 \mu\text{g ml}^{-1}$). The response of spleen cells from control rats (NR) to Con A (a), PHA (b) or LPS (c) was assayed at the same time as for rats born to microfilaraemic mothers (MR). [$\text{Me-}^3\text{H}$]thymidine incorporation was measured as in Fig. 1. a, b, and c are the mean $\pm$ one standard deviation stimulation indices for normal spleen cells in response to Con A, PHA and LPS respectively.

Further experiments were undertaken to examine whether the transfer of microfilariae from mothers to the litter occurs through the placenta during gestation or via the milk during lactation. The litters born to microfilaraemic mothers were removed immediately after birth and kept with non-infected foster mothers for suckling. They always showed the presence of microfilariae in their circulation. In contrast, microfilariae were never observed in any of the litters born to non-infected mothers and kept with microfilaraemic foster mothers for nursing. The body fluids from the fetus removed from microfilaraemic mothers by caesarian delivery always contained microfilariae. These data strongly suggest that the transfer of microfilariae occurs through the placenta and not through the milk.

The possible tolerizing effect of the transplacentally transferred microfilariae was tested by measuring the immune response of the offspring towards filarial antigen *in vitro* and *in vivo*. Antigens prepared from *D. viteae* adult worms had no effect

on spleen cells from normal Fischer rats *in vitro*. They were able to stimulate cells from rats in which transplacental transfer of microfilariae had occurred but only when the microfilariae had disappeared from the peripheral blood (Fig. 1). There was no significant response of spleen cells from normal or infected rats towards antigens of *Ascaris lumbricoides* at any time following birth. The spleen cells from normal or microfilariae infected rats immunized with ovalbumin (OA, a total of $200 \mu\text{g}$ with complete Freund's adjuvant per rat) at day 70 after birth were, however, stimulated in the presence of OA (data not shown). In the studies with bancroftian⁸ and brugian⁹ filariases or with various experimental filariases¹⁰⁻¹² in which both adult worms and microfilariae coexist, it has been shown that the infected individuals or animals with microfilaraemia respond poorly or not at all to filarial antigens. The results of the present study indicate that the causative agent for the antigen-specific cellular unresponsiveness towards *D. viteae* antigens may be due to microfilariae or their derived factors.

The response of spleen cells, obtained from rats born to microfilaraemic mothers and showing the presence of microfilariae in the circulation, to phytohaemagglutinin (PHA) and concanavalin A (Con-A) was depressed but their response to lipopolysaccharide (LPS) was unaffected (Fig. 2). This suggests that the function of the T lymphocytes that respond to PHA and Con-A was impaired. However the significance of the depressed response of spleen cells to mitogens is controversial, although it has often been observed in diseases where defects in cellular immunity (antigen specific or nonspecific) have been documented¹³.

Antibodies and soluble filarial antigens may be transferred from mother to fetus across the placenta. We were, however, unable to detect a significant level of *D. viteae* circulating antigens in the serum obtained from rats born to microfilaraemic mothers at any period after birth using anti-*D. viteae* hyperimmune rabbit serum in the radioimmunoassay PEG assay (data not shown). This could have been due to the dilution of circulating antigens (present in the mothers at a relatively low level) after transfer. Alternatively, it may be that the numbers of microfilariae present in the offspring were too low to produce detectable circulating antigen or that the adult worm is the only source of such antigen. Antibodies to *D. viteae* filarial antigens or to microfilariae were not detected by haemagglutination or the immunofluorescence assay in the serum from the offspring as long as microfilariae circulated in their peripheral blood (data not shown). The results show that although the response of B lymphocytes to LPS was not affected, the production of parasite specific antibodies was impaired.

The effects of acquired tolerance on the susceptibility or resistance to *D. viteae* infection were also investigated. The

Table 1 Microfilaraemia and adult worm recovery on various days after infection or birth

Infection	Mean microfilaraemia $\pm$ s.d. per 100 mm ³ on various days after infection										Mean no. of adults recovered from 3 rats on day	
	7	20	30	40	60	70	90	110	120	130	45	60
Female rats implanted with 10 <i>D. viteae</i> female worms	90 $\pm$ 40	680 $\pm$ 210	650 $\pm$ 140	210 $\pm$ 80	240 $\pm$ 72	190 $\pm$ 56	70 $\pm$ 32	40 $\pm$ 20	10 $\pm$ 6	0	5 $\pm$ 3*	ND
Rats born to microfilaraemic	20 $\pm$ 8	24 $\pm$ 10	18 $\pm$ 6	16 $\pm$ 6	18 $\pm$ 8	16 $\pm$ 4	16 $\pm$ 6	10 $\pm$ 4	4 $\pm$ 2	0	—	—
Rats born to noninfected mothers	0	0	0	0	0	0	0	0	0	0	—	—
MR infected with <i>D. viteae</i> stage-3 larvae†	18 $\pm$ 6	18 $\pm$ 6	16 $\pm$ 4	18 $\pm$ 8	24 $\pm$ 6	60 $\pm$ 12	70 $\pm$ 10	20 $\pm$ 8	4 $\pm$ 2	0	7 $\pm$ 3‡	4 $\pm$ 2‡
NR infected with <i>D. viteae</i> stage-3 larvae†	0	0	0	0	0	0	0	0	0	0	0	0

Samples of blood in duplicate (10 or 20 μl) were taken with a glass disposable microsampling pipette from the retroorbital plexus of rat, spread directly onto a slide, covered with a cover slip and immediately examined under the microscope and counted. As the number of microfilariae decreased, blood samples were treated with distilled water and microfilariae were isolated on 5 μm Millipore membranes. Microfilariae were then resuspended in Hanks Wallace solution and counted. Rats born either to microfilaraemic or non-infected mothers were infected at day 75 after birth with 100 *D. viteae* stage-3 larvae. Ticks (*Ornithodoros tartakowskyi*) infected 4 weeks previously were washed briefly in 70% alcohol, rinsed in sterile saline and then dissected with fine sharpened needles in Ringer's solution containing antibiotics. The stage-3 larvae obtained were washed free of tick tissues, counted and injected subcutaneously into rats. Five rats per group were used. ND, not determined.

* Worms were calcified.

† Infection rate with *D. viteae* was 80%. This includes 3 $\pm$ 2 which were dead and encapsulated in a granulomatous nodule.

‡ This includes 2 $\pm$ 1 which were dead and encapsulated in a granulomatous nodule.

male rats born to microfilaraemic mothers and harbouring microfilariae in their circulation became susceptible to *D. viteae*. A varying percentage of infective larvae of *D. viteae* became established and developed into adult worms. These worms produced microfilariae which were observed in the peripheral blood of the rats from about day 55 after infection. In contrast, no microfilariae were seen in the peripheral blood and no worms were found at post-mortem in rats, age-matched and born to non-infected mothers, following inoculation with *D. viteae* infective larvae 75 days after birth (Table 1). On the other hand, when rats born either to microfilaraemic or to non-infected mothers were exposed to *Schistosoma mansoni* cercariae there were no differences between the numbers of *S. mansoni* worms (immature) recovered at day 21 after infection from each group (data not shown). These results suggest that a state of antigen-specific immune unresponsiveness had developed in the animals as a result of prior sensitization with specific parasite antigens in the early stages of their life. In this investigation, we have shown that the suppression of the host's immune responsiveness also suppresses innate resistance and permits a filarial nematode to establish itself in an otherwise refractory host.

We thank Bernadette Bonnel, Jean-Paul Lejeune, Anita Caron and Pierre Billaut for technical assistance and maintenance of the animal colony. This work was supported by CNRS (ERA 422) and INSERM(U 167).

Received 20 April; accepted 12 July 1982.

- Mantovani, A. *Parasitologia* 6, 1-3 (1964).
- Bloomfield, R. D., Suarez, J. R. & Malangit, A. C. *J. natn. med. Ass.* 70, 587-598 (1978).
- Sucharit, S. & Rongsriyarn, Y. S. E. *Asian J. trop. Med. publ. Hlth* 11, 416 (1980).
- Dresser, D. & Mitchison, N. *Adv. Immun.* 8, 129-181 (1968).
- Etlinger, H. & Chiller, J. *J. Immun.* 122, 2558-2563 (1979).
- Hasek, M. & Chutna, J. *Immun. Rev.* 46, 3-26 (1979).
- Haque, A. et al. *Clin. exp. Immun.* 43, 1-9 (1981).
- Ottesen, E. A., Weller, P. F. & Heck, L. *Immunology* 33, 413-421 (1977).
- Piessens, W. F. et al. *J. clin. Invest.* 65, 172-179 (1980).
- Portaro, J. K., Britton, S. & Ash, L. R. *Expl Parasit.* 40, 438-446 (1976).
- Weller, P. F. *Cell Immun.* 37, 369-382 (1978).
- Haque, A. A., Ogilvie, B. M. & Capron, A. *Expl Parasit.* 52, 25-34 (1981).
- Bryceson, A. in *Prog. Immun.* 2, 161-170 (1974).

Lithium chloride induces partial responsiveness to LPS in nonresponder B cells

Shigeaki Ishizaka & Göran Möller

Department of Immunobiology, Karolinska Institute, Wallenberg Laboratory, Lilla Frescati 7, S-104 05 Stockholm 50, Sweden

The lipopolysaccharide (LPS)-nonresponder mouse strains C3H/HeJ, C57BL/10ScCR and C57BL/10ScN do not respond to LPS acting as a polyclonal B-cell activator^{1,2}, a mitogen¹⁻⁵, or an adjuvant^{6,7}. The genetic basis for the defective LPS responses has been extensively studied in C3H/HeJ and C57BL/10ScCR mice, in which it was demonstrated that a single gene locus on chromosome 4 was responsible for LPS unresponsiveness^{8,9}. Lithium chloride, a potent inhibitor of adenylate cyclase, not only improved lymphocyte activity in a patient with adenosine deaminase deficiency^{10,11} but also enhanced the phytohaemagglutinin (PHA)-induced responses of normal human lymphocytes^{12,13}. Therefore, we investigated whether LiCl could restore LPS responsiveness in spleen cells of C3H/HeJ mice. We show here that LPS, in the presence of LiCl, induced polyclonal IgM and IgG antibody formation and DNA synthesis in C3H/HeJ mouse spleen cells *in vitro*. Moreover, LiCl (10 mM), which by itself is non-mitogenic, increased RNA synthesis in spleen cells from both LPS-nonresponder and high responder strains; in contrast, LPS failed to increase RNA synthesis in cells from such LPS-nonresponder strains as C3H/HeJ and B10ScCR mice.

C3H/HeJ mouse spleen cells (10^7 cells ml^{-1}) were cultured with LPS in the presence of (10 mM LiCl) to determine whether LiCl affects the B-cell nonresponsiveness to LPS in these cells.

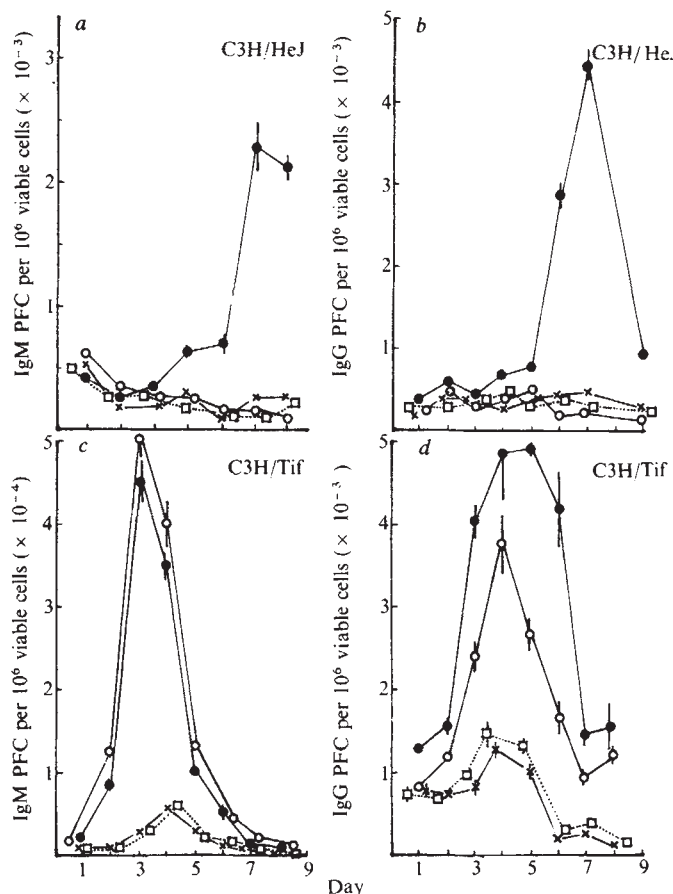


Fig. 1 Kinetics of polyclonal IgM and IgG antibody production induced by LPS in the presence of LiCl in C3H/HeJ and C3H/Tif spleen cells. Spleen cells (10^7 cells ml^{-1}) from unprimed C3H/HeJ or C3H/Tif mice were cultured for 9 days with $150 \mu\text{g ml}^{-1}$ of LPS extracted from *Escherichia coli* 055:B5 by the phenol-water method (given by T. Holme, Karolinska Institute) in the presence or absence of LiCl (10 mM) in RPMI 1640 medium (Gibco) plus 10% fetal bovine serum. IgM and IgG PFC per 10^6 viable cells were counted after the indicated days by using the protein A plaque assay²⁴. a, b: IgM and IgG PFC in C3H/HeJ cells induced by $150 \mu\text{g ml}^{-1}$ LPS + 10 mM LiCl (●), $150 \mu\text{g ml}^{-1}$ LPS (○), or 10 mM LiCl (□); ×, unstimulated control. c, d: IgM and IgG PFC in C3H/Tif cells induced by $150 \mu\text{g ml}^{-1}$ LPS + 10 mM LiCl (●), $150 \mu\text{g ml}^{-1}$ LPS (○) or 10 mM LiCl (□); ×, unstimulated control. Each point represents the mean $\pm$ s.e. (vertical bars) of triplicate experiments.

(Preliminary tests suggested that 10 mM is the optimal concentration of LiCl.) In the presence, but not the absence, of LiCl, LPS induced polyclonal IgM and IgG antibody responses in the spleen cells that peaked after 6-7 days of culture. LiCl alone did not stimulate polyclonal IgM and IgG antibody responses (Fig. 1), and was non-toxic to cultured spleen cells, as judged by Trypan blue exclusion. On the other hand, LPS induced a polyclonal IgM response in spleen cells from high responder strain C3H/Tif that peaked on day 3. LPS plus LiCl did not affect the magnitude of the IgM response, whereas the IgG response was increased compared with that induced by LPS alone (Fig. 1c, d). Thus, the magnitude of the observed stimulation by LPS plus LiCl in spleen cells from C3H/HeJ mice (indices of maximal stimulation of IgM and IgG were 14.2 and 9.4 plaque-forming cells (PFC) respectively) was very similar to that induced by LPS alone in C3H/Tif spleen cells (maximal stimulation indices of IgM and IgG were 17.5 and 7.9 respectively), because the background PFC of high responder cells was higher than that of nonresponder cells (Fig. 1). LiCl plus LPS also stimulated DNA synthesis in C3H/HeJ and C3H/Tif spleen cells, although the response was smaller with C3H/HeJ cells. Neither LiCl nor LPS alone stimulated DNA synthesis in C3H/HeJ spleen cells (Table 1).

Table 1 Effect of LiCl on LPS-induced DNA synthesis in C3H/HeJ and C3H/Tif spleen cells

Strain	Addition to culture	1	2	c.p.m. on day: 3	4	5
C3H/HeJ	None	402 ± 26	281 ± 14	200 ± 87	185 ± 82	226 ± 82
	LPS	570 ± 63	242 ± 13	311 ± 34	304 ± 57	347 ± 22
	LiCl	531 ± 61	147 ± 6	146 ± 36	142 ± 12	245 ± 52
	LiCl + LPS	662 ± 62	580 ± 19	1,052 ± 175	748 ± 58	800 ± 107
C3H/Tif	None	186 ± 33	167 ± 18	307 ± 50	902 ± 115	999 ± 35
	LPS	407 ± 86	2,432 ± 248	5,338 ± 261	10,897 ± 815	5,969 ± 224
	LiCl	253 ± 55	192 ± 52	369 ± 30	1,579 ± 223	1,167 ± 258
	LiCl + LPS	366 ± 51	5,387 ± 492	11,441 ± 370	17,494 ± 745	7,606 ± 232

C3H/HeJ spleen cells (10^6 cells ml^{-1}) were cultured in RPMI 1640 medium containing 5% fetal bovine serum. LPS was added to a final concentration of $30 \mu\text{g ml}^{-1}$. LiCl was also added to a final concentration of 10 mM per well of a round-bottomed microtitre plate (Nunc). At appropriate times, the wells were then pulsed with $1 \mu\text{Ci ml}^{-1}$ of ^3H -thymidine for 6 h. The cells were collected using a multiple automated sample collector and processed for scintillation counting. Results are the mean c.p.m. $\pm$ s.e. of triplicate experimental cultures.

Polyclonal IgM antibody responses induced by LPS in cells from high responder mice peaked at day 3 of culture in the conditions used, whereas the peak response of cells from LPS-nonresponder mice occurred at day 6 (Fig. 1). There was a kinetic difference between LPS-high responder and nonrespon-

der mice with respect to LPS-induced polyclonal IgM responses. It seems unlikely that the polyclonal IgM responses induced by LPS in spleen cells from C3H/HeJ mice was secondary to polyclonal T-cell activation induced by LiCl, as LiCl did not stimulate DNA or RNA synthesis in T cells and thymocytes from C3H/HeJ mice (data not shown). Furthermore, LPS non-responsiveness in B cells from C3H/HeJ mice is due to a B-cell defect rather than the activation of suppressor T cells or the dysfunction of helper T cells and macrophages^{14,15}. It is possible that LiCl induces the expression of a LPS receptor in B cells from C3H/HeJ mice, but it is unknown whether there are any differences in the binding of LPS to B cells from LPS-high responder and nonresponder mice^{16,17}. Therefore, the possibility of B-cell activation in LPS-nonresponder mice induced by this mechanism remains open.

LiCl did not induce DNA synthesis or IgM and IgG antibody synthesis in cells from C3H/HeJ mice, but increased ^3H -uridine incorporation in cells from C3H/HeJ mice; LPS did not have this effect. LPS plus LiCl increased DNA, RNA and IgM antibody synthesis in C3H/HeJ spleen cells and B cells (Table 2), suggesting that nonresponsiveness to LPS in C3H/HeJ spleen cells was partly overcome by the new RNA synthesis induced by LiCl. LiCl produced greater stimulation of RNA synthesis in cells from the LPS-nonresponder strains C3H/HeJ and B10ScCR than in cells from the LPS-high responder strains C3H/Tif and B10. Furthermore, LPS plus LiCl induced later RNA synthesis in cells from LPS-nonresponder mice than in cells from high responder mice (Fig. 2). These results show that the increase in RNA synthesis and polyclonal antibody synthesis induced by LPS plus LiCl occurs later in C3H/HeJ cells than in cells from a LPS-high responder strain.

LPS induced a relatively low rate of RNA synthesis in LPS-high responder strains compared with DNA and IgM synthesis, in contrast to results obtained with other polyclonal B-cell activators¹⁸. Furthermore, recent cell fusion studies indicate that LPS nonresponsiveness of C3H/HeJ cells is due to a dysfunction of the cytoplasm and not the nucleus¹⁹. Therefore, it seems possible that LPS stimulates translation of mRNA rather than new mRNA synthesis. Although LPS failed to induce RNA synthesis in spleen cells of LPS-nonresponder mice, the nonresponsiveness to LPS is probably alleviated by direct or indirect effects of LiCl on murine chromosomes^{20,21} leading to induction of new RNA synthesis, perhaps through the modification of cyclic AMP-dependent events in cells. In addition, there is indirect evidence that α -amanitin, an inhibitor of RNA synthesis, inhibited polyclonal antibody responses induced by κ -carrageenan²², but not LPS-induced responses at an early stage of the response. Also, actinomycin D, an inhibitor of ribosomal RNA synthesis, suppressed polyclonal antibody formation induced by both LPS and κ -carrageenan¹⁸. This suggests that LPS stimulates the translation of RNA by ribosomes in B cells of LPS-nonresponder mice and that LiCl may

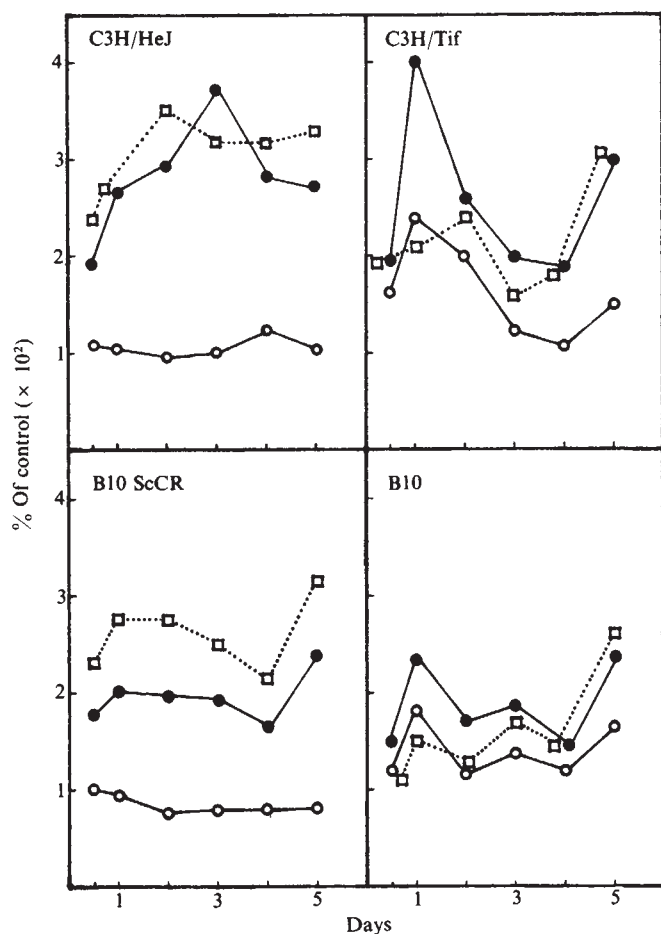


Fig. 2 Time course of RNA synthesis in spleen cells from LPS-nonresponder and high responder mice induced by LPS plus LiCl. Spleen cells (2×10^6 cells per 0.2 ml) from C3H/HeJ, B10ScCR, C3H/Tif and B10 mice were cultured for 5 days with $150 \mu\text{g ml}^{-1}$ LPS, 10 mM LiCl, or $150 \mu\text{g ml}^{-1}$ LPS plus 10 mM LiCl per well of microplates (Nunc) in RPMI 1640 medium supplemented with 5% fetal bovine serum. These cell cultures were pulsed with $0.2 \mu\text{Ci } ^3\text{H}$ -uridine (specific activity 37 Ci mmol^{-1} ; NEN) for 3 h at various times. The reactions were stopped by adding cold 5% trichloroacetic acid, and cells were collected with a multisample collector for analysis. RNA synthesis induced by LPS (○), LiCl (□), or LPS plus LiCl (●) are expressed as the mean experimental c.p.m. per control c.p.m. $\times 10^{-2}$.

Table 2 DNA, RNA and IgM synthesis in C3H/HeJ B cells induced by LPS and LiCl

Addition	C3H/HeJ spleen cells		C3H/HeJ B cells		
	DNA (c.p.m.)	RNA (c.p.m.)	DNA (c.p.m.)	RNA (c.p.m.)	IgM PFC per 10^6 viable cells
None	564 ± 109	189 ± 11	356 ± 26	473 ± 63	170 ± 15
LPS	625 ± 59 (1.1)	207 ± 8 (1.1)	485 ± 36 (1.4)	260 ± 22 (0.5)	180 ± 20 (1.1)
LiCl	571 ± 133 (1.0)	753 ± 88 (4.0)	268 ± 33 (0.8)	2,721 ± 197 (5.8)	202 ± 15 (1.2)
LPS + LiCl	2,019 ± 247 (3.6)	790 ± 207 (4.2)	805 ± 69 (2.3)	2,328 ± 204 (4.9)	513 ± 11 (3.0)
Con A	42,786 ± 1,342 (75.9)	NT	498 ± 22 (1.4)	NT	NT

C3H/HeJ B-cell suspensions were prepared by treatment of spleen cells with monoclonal anti-Thy 1.2 antibody (final concentration 10^{-3} ; NEN) and complement²³, followed by removal of adherent cells in plastic Petri dishes as described previously²². C3H/HeJ B cells (2×10^6 cells per 0.2 ml) were cultured with $150 \mu\text{g ml}^{-1}$ LPS, 10 mM LiCl, $150 \mu\text{g ml}^{-1}$ LPS plus 10 mM LiCl or concanavalin A (Con A) in RPMI 1640 medium supplemented with 5% fetal bovine serum for 3 days. Synthesis of DNA and RNA was measured by pulsing with $1 \mu\text{Ci ml}^{-1}$ ^3H -thymidine and $1 \mu\text{Ci ml}^{-1}$ ^3H -uridine for 3 h. IgM PFC were estimated by using the protein A plaque assay²⁴. The values in parentheses are stimulation indexes. NT, not tested.

induce the production of ribosomal RNA rather than mRNA and then elicit polyclonal antibody synthesis. In LPS-nonresponder mice, LPS could not efficiently stimulate translation due to insufficient polyribosome formation, without the produc-

tion of new ribosomal RNA induced by LPS, thereby resulting in nonresponsiveness to LPS. The alteration in LPS-nonresponder cells by LiCl may be important in the analysis of the biochemical events of B-cell activation by LPS.

Received 25 March; accepted 20 July 1982.

- Watson, J. & Riblet, R. *J. exp. Med.* **140**, 1147–1161 (1974).
- Coutinho, A., Forni, L., Melchers, F. & Watanabe, T. *Eur. J. Immun.* **7**, 325–328 (1977).
- Sultzer, B. M. *Infect. Immunity* **5**, 107–113 (1972).
- Vogel, S. N., Hansen, C. T. & Rosenstreich, D. L. *J. Immun.* **122**, 619–622 (1979).
- Coutinho, A., Gronowicz, E. & Sultzer, B. M. *Scand. J. Immun.* **4**, 139–143 (1975).
- Skidmore, B. J., Chiller, J. M., Morrison, D. S. & Weigle, W. O. *J. Immun.* **114**, 770–775 (1975).
- Talcott, J. A., Ness, D. B. & Grumet, F. C. *Immunogenetics* **2**, 507–515 (1975).
- Watson, J., Riblet, R. & Taylor, B. A. *J. Immun.* **118**, 2088–2093 (1977).
- Coutinho, A. & Meo, T. *Immunogenetics* **7**, 17–24 (1978).
- Polmar, S. H. in *Inborn Errors of Immunity and Phagocytosis* (eds Guttie, F., Seakins, J. W. T. & Harkness, R. A.) (MTP Press, Lancaster, 1979).

- Gelfand, E. W. & Dush, H.-M. in *In Vitro Induction and Measurement of Antibody Synthesis in Man* (eds Fauci, A. S. & Ballieux, R. E.) 309 (Academic, New York, 1979).
- Gelfand, E. W., Dush, H.-M., Hastings, D. & Shore, A. *Science* **203**, 365–367 (1979).
- Shenkman, L., Borkowsky, W. & Shopsin, B. *Med. Hypotheses* **6**, 1–6 (1980).
- Glode, L. M., Scher, L., Osborne, B. & Kern, M. J. *Immun.* **125**, 102–107 (1980).
- Coutinho, A. *Scand. J. Immun.* **5**, 129–140 (1976).
- Gregory, S. H., Zimmerman, D. H. & Kern, M. J. *Immun.* **125**, 102–107 (1980).
- Forni, L. & Coutinho, A. *Eur. J. Immun.* **8**, 56–62 (1978).
- Ishizaka, S. & Morisawa, S. (in preparation).
- Watanabe, T. & Ohara, J. *Nature* **290**, 58–59 (1981).
- Szabo, K. T. *Lancet* **ii**, 849 (1969).
- Szabo, K. T. *Nature* **225**, 73–75 (1970).
- Ishizaka, S., Otani, S. & Morisawa, S. *J. Immun.* **118**, 1213–1218 (1977).
- Gorczynski, R. M., Miller, R. G. & Phillips, R. A. *J. Immun.* **108**, 547–551 (1972).
- Gronowicz, E., Coutinho, A. & Melchers, F. *Eur. J. Immun.* **6**, 588–590 (1976).

Random components in mutagenesis

P. L. Foster*, E. Eisenstadt†‡ & J. Cairns†

* Interdisciplinary Programs in Health,

† Department of Microbiology and ‡ Laboratory of Toxicology, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

The mutability of DNA varies enormously from one base pair to another. Part of this variation is due to the specificity of the reaction between mutagen and base, but much of the variation is due to unknown causes. A genetic system developed by Miller and colleagues¹ allows the mutation frequencies of a large number of different base pairs in the *lacI* gene of *Escherichia coli* to be compared. For example, Coulondre and Miller² found that the sites most readily mutated by UV light are almost 100 times more often mutated than the least susceptible sites. A recently completed study³ of mutagenesis with neocarzinostatin (NCS) in the *lacI* gene has prompted us to re-examine some previous studies^{2,4,5} of mutagenesis in this gene. Our analysis, reported here, suggests that the mutations induced by certain mutagens fall into two classes: mutations in one class are clearly distributed non-randomly, that is, they are very common at some sites and significantly less common at others; mutations in the second class, however, occur at low frequency and appear to be randomly distributed. Both classes of mutations seem to occur only at damaged bases.

One of the largest collections of fully characterized point mutations is a set of about 600 nonsense mutations generated in the *lacI* gene by UV light². There are 64 sites in this gene that can yield amber (TAG) or ochre (TAA) codons by single base changes. Overall, the distribution of UV-induced mutations among these sites is obviously non-random; 3 of the 64 possible nonsense codons were each represented more than 50 times, 3 did not appear in the collection at all and many occurred

only once or twice. Because UV-mutagenesis clearly involves non-random elements, it has often been assumed that the entire process is non-random and that each base pair occupies some unique position in a continuum of mutability. However, if we disregard both the base substitutions being monitored and the distribution of events over the genetic map, and instead simply look at the distribution of frequencies, it becomes apparent that the rarer mutations form a Poisson distribution (that is, their frequencies appear to be randomly distributed) with a mean of 2.7 occurrences per site (Fig. 1a). Of the 64 sites, 41 fall into this distribution. (We have included in this set the 3 codons in which no nonsense mutations were detected, because a mean occupancy of 2.7 would be expected to leave 3 out of 41 sites unoccupied.)

Solely on the basis of the frequencies of the different mutations, we therefore conclude that UV-induced mutations arise from two classes of events: apparently random, low-frequency occurrences (LFOs) that account for about one-third of the mutations, and non-random, high-frequency occurrences (HFOs) that account for the rest. This conclusion is greatly strengthened when we consider the base sequences at the 64 sites. As pointed out by Coulondre *et al.*⁵, all of the dozen or so 'hotspots' for UV-induced G·C to A·T transitions found in the *lacI* gene are at sites of adjacent pyrimidines. Based on our analysis, we find that all 23 of the sites for HFOs are base pairs at which the pyrimidine has a pyrimidine next to it; in contrast, 15 of the 41 LFO sites involve base pairs where the pyrimidine has purines on both sides of it. This difference between the sites for HFOs and LFOs is highly significant ($\chi^2 = 9.4$, $P < 0.005$). Thus, it seems likely that UV-induced HFOs require interaction between adjacent pyrimidines (for example, they may be due to pyrimidine dimers⁶ or some other lesion involving a pyrimidine pair⁷) whereas LFOs do not. As the UV-induced LFOs occur at every monitorable site in the *lacI* gene and can generate each of the five base pair changes that yield nonsense codons (G·C to A·T, T·A or C·G; A·T to T·A or C·G), it seems likely that UV light is producing, at

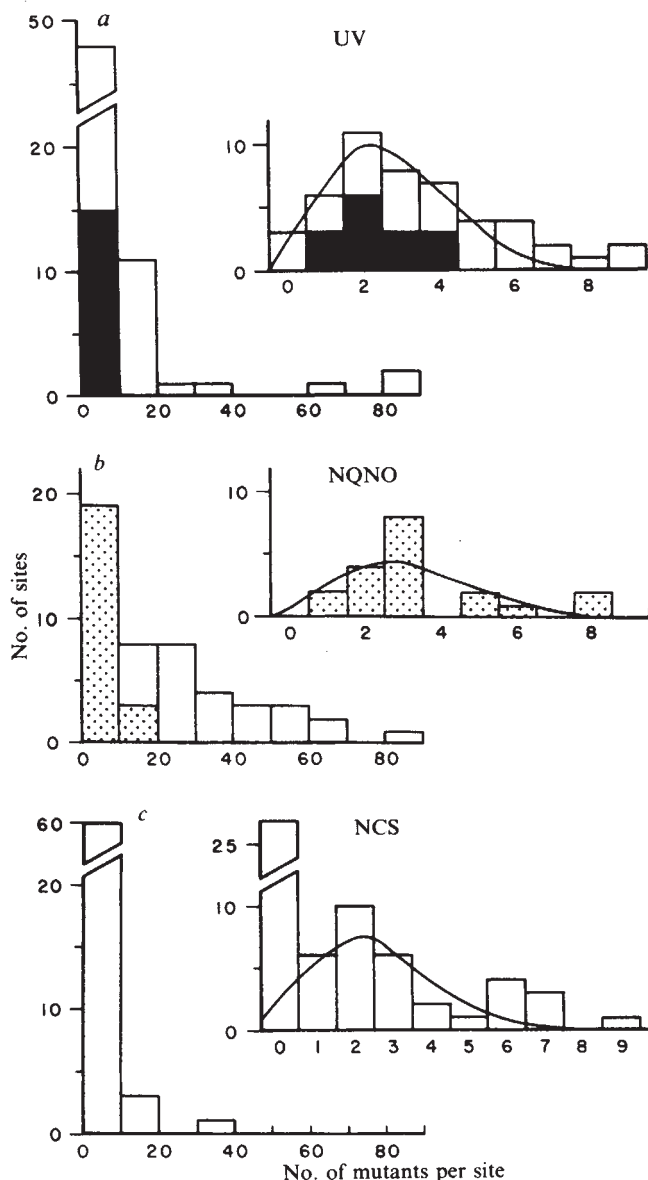


Fig. 1 The frequency distributions of approximately 600 UV-induced² (a), 1,000 NQNO-induced³ (b) and 200 NCS-induced³ (c) amber and ochre mutations in the *lacI* gene of *E. coli*. The histograms show the number of sites at which k mutations were found. The inset histograms show the lower end of the distributions for individual values of k from 0 to 9. These have been matched to Poisson distributions, whose means were estimated using the relationship:

$$m = (k+1) \frac{N_{k+1}}{N_k}$$

where N_k and N_{k+1} are the number of sites at which occurred k and $k+1$ mutations, respectively. The observed values of N_k for several consecutive classes were used to calculate m . The total number of sites (N_T) expected to fall into the Poisson distribution was calculated from the observed values of N_k using the relationship:

$$N_T = \frac{\sum_{k=1}^5 N_k}{e^{-m} \sum_{k=1}^5 m^k / k!}$$

The total number of mutations (T) in the Poisson distribution was then calculated as:

$$T = mN_T$$

a, UV-induced mutations at all sites are shown; sites where the pyrimidine of the target base pair is not adjacent to another pyrimidine are shaded. The Poisson distribution includes 111 mutations at 41 sites ($N_k \leq 6$); $m = 2.7$ was estimated from N_0 to N_2 . b, NQNO-induced mutations at G-C sites only are shown; sites where G-C transversions occur are stippled. The Poisson distribution included 63 mutations at 18 sites ($N_k < 8$); $m = 3.3$ was estimated from N_1 and N_2 . c, NCS-induced mutations at all sites are shown. The Poisson distribution included 72 mutations at 29 sites ($N_k \leq 6$); $m = 2.5$ was estimated from N_1 and N_2 .

random, all possible base substitutions at every base pair.

For a base pair to be a site for UV-induced HFOs, it clearly is not enough that its pyrimidine be adjacent to another pyrimidine. Of the 49 base pairs in the *lacI* gene that involve adjacent pyrimidines and are capable of generating nonsense codons, only 23 were sites for HFOs. For example, of the 14 Py-C-A-G and Py-C-A-A sequences in the *lacI* gene where transition of the C to a T yields a nonsense codon, 10 are sites for HFOs and cover a 10-fold range of mutability; the remaining 4 are sites for LFOs only. So some factor other than neighbouring base sequence must be important in determining the frequency of mutation at these sites. As the loss of the *uvrB*-dependent excision repair pathway has no effect on the relative UV mutability of most sites⁴, accessibility to this repair pathway cannot account for most of the variation in mutation frequency. We conclude, therefore, that the mutability of different pairs of adjacent pyrimidines is probably determined by the effect that some factor, such as DNA secondary structure, has on the frequency of formation of premutational lesions.

Mutagenesis by UV light requires induction of the 'SOS response'⁸. Because LFOs seem to be random and indiscriminate, it was conceivable that they represent untargeted mutagenesis due to error-prone DNA repair or replication. We therefore extended our analysis to the published results for 4-nitroquinoline-*N*-oxide (NQNO), which is also an SOS-dependent mutagen⁹. Coulondre and Miller² have shown that NQNO is highly specific for G-C base pairs, with a strong preference for producing G-C to A-T transitions. We see here that NQNO, like UV light, produces a non-random distribution of HFOs plus an apparently random distribution of LFOs (Fig. 1b). However, in contrast to the UV results, the mutations that are produced by the NQNO-induced HFOs and LFOs are distinguishable—all the G-C transitions arise from HFOs, whereas nearly all the G-C transversions arise from LFOs (any G-C transitions produced by LFOs cannot be detected because the sites for transitions are dominated by HFOs). Since untargeted mutagenesis would be expected to include base substitutions of A-T sites, the failure of NQNO to mutate A-T base pairs implies that both HFOs and LFOs result from targeted events. We can conclude, therefore, that the low level of apparently random mutagenesis following treatment with mutagens such as NQNO or UV light cannot be accounted for by untargeted mutagenesis.

Another SOS-dependent mutagen that we have examined is NCS (Fig. 1c). Like UV light and NQNO, it produces both HFOs and LFOs; the LFOs include examples of each of the five kinds of base changes that can be detected as nonsense mutations. However, NCS mutates only 37 of the 64 sites in the *lacI* gene; the sites it fails to mutate include examples of all monitorable base changes³. As in the case of NQNO, the failure of NCS to mutate nearly half the monitorable sites in the *lacI* gene would be difficult to explain if untargeted mutagenesis were making a major contribution to SOS-dependent mutagenesis.

Although we have looked at only a few mutagens, we now realize that the available data are less informative than they seem at first sight. For example, because of the nature of the genetic code, an analysis of nonsense mutations cannot distinguish which of the four pyrimidine pairs (T-T, T-C, C-T, or C-C) is the best target for UV mutagenesis; any effect attributable to the pyrimidine pair is confounded by the constraints on both the event being monitored (which can be transitions or transversions at cytosines, but only transversions at thymines) and the reading frame (for example, substitutions for thymine cannot be monitored at the first position of a codon). To separate these factors would require an analysis of missense mutations, which are not subject to the same base sequence restrictions. But there is yet another problem. In any mutant collection, the rarer events (LFOs) will be present in rather small numbers and, because of this, their frequency distribution will be dominated by the variance of the Poisson distribution. Thus, although we have said that LFOs appear to be randomly

distributed, this suggestion needs to be confirmed by a much larger collection of LFOs.

We can, however, draw some important conclusions about LFOs. First, like HFOs, they are likely to be targeted, that is, to occur only at damaged bases. Second, they can account for a significant proportion of the mutations arising after exposure to a mutagen (for example, one-third of the UV-induced mutations). Last, they seem to be able to induce all base changes—the LFO lesion seems to be non-informational and to allow the insertion of a base at random. These conclusions suggest that, unlike HFO lesions, which are clearly specific to each mutagen, the LFO lesion may be the same for a number of mutagens; one example of such a common lesion would be an apurinic or apyrimidinic site.

P.L.F. was a fellow of the Interdisciplinary Programs in Health and was supported by grant CR807809-01-1 from the US Environmental Protection Agency. E.E. was supported in part by US NIH grant CA26135.

Received 12 April; accepted 26 July 1982.

1. Miller, J. H. in *The Operon* (eds Miller, J. H. & Reznikoff, W. S.) 31–88 (Cold Spring Harbor Laboratory, New York, 1978).
2. Coulondre, C. & Miller, J. H. *J. molec. Biol.* **117**, 557–606 (1977).
3. Foster, P. L. & Eisenstadt, E. *J. Bact.* (submitted).
4. Todd, P. A. & Glickman, B. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4123–4127 (1982).
5. Coulondre, C., Miller, J. H., Farabaugh, P. J. & Gilbert, W. *Nature* **274**, 775–780 (1978).
6. Fisher, G. J. & Johns, H. E. in *Photochemistry and Photobiology of Nucleic Acids* Vol. 2 (ed. Wang, S.-Y.) 225–295 (Academic, New York, 1976).
7. Lippke, J. A., Gordon, L. K., Brash, D. E. & Haseltine, W. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3388–3392 (1981).
8. Witkin, E. M. *Bact. Rev.* **40**, 869–907 (1976).
9. Kondo, S., Ichikawa, H., Iwa, K. & Kato, T. *Genetics* **66**, 187–217 (1970).

Unique cell lines harbouring both Epstein-Barr virus and adult T-cell leukaemia virus, established from leukaemia patients

Naoki Yamamoto, Tadashi Matsumoto*, Yoshio Koyanagi, Yuetsu Tanaka & Yorio Hinuma

Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606, Japan

* Institute of Cancer Research, Faculty of Medicine, Kagoshima University, Kagoshima 880, Japan

Members of three distinct classes of animal virus have been associated with naturally occurring neoplasias in man: Epstein-Barr virus (EBV)¹, a DNA virus belonging to herpesvirus group, papillomavirus², and a novel human RNA (retro) virus^{3–8}, human T-cell leukaemia virus or adult T-cell leukaemia (ATL) virus. We have established seven continuous cell lines from ATL patients and 0.1–7% of these cells consistently express ATL-specific antigens (ATLA)^{9,10}. EBV-associated nuclear antigen (EBNA)¹⁰ is also found in more than 90% of these cells. We have cloned cells from two of these lines and show here that both ATLA and EBNA were present in the same B-cell clone carrying surface immunoglobulin (sIg).

Patients with ATL investigated in this study were admitted from September 1980 to December 1981 to the hospitals in the Kagoshima Prefecture, where ATL is endemic¹¹. Clinical features and some of the data from laboratory investigations of the patients are shown in Table 1. Most frequently they reveal leukocytosis, with a high percentage of leukaemic cells showing indented or lobulated nuclei. Most of these cells spontaneously form rosettes with sheep erythrocytes (E). The percentage of cells with sIg is usually very low. All patients tested had antibodies against EBV and ATL.

Peripheral leukocytes separated by Ficoll-Conray gradient centrifugation were cultured in 10-cm plastic tubes (1 ml per tube) with RPMI 1640 medium supplemented with 10–20% fetal calf serum and antibiotics. The initial cell density was adjusted to 5×10^6 cells per ml. The cultures were incubated at 37 °C in a humidified 7.5% CO₂ atmosphere for 1 week without any change of medium. Thereafter, they were fed twice weekly. After 1 month of culture, they started to grow continuously and the number of cells increased rapidly. The originating cell lines were designated ATL 1–7. They grew in suspension forming clumps of cells and have been maintained in continuous culture for 6–22 months. Morphologically, they were lymphoblastoid cell lines with a doubling time of about 24–48 h. The cells unexpectedly had sIg in 5–75% of the population (sIgG + sIgM) and were mostly negative for E-receptors as shown in Table 2. No T-cell growth factor¹² was used throughout these experiments. The establishment and more detailed characteristics of these lines will be reported elsewhere.

The acetone or methanol fixed cells were examined for EBNA and for ATLA by indirect immunofluorescence as described previously^{7–10}. For EBNA tests, standard sera obtained from healthy persons were used. For ATLA staining, human sera from ATL patients were used initially but it was impossible to detect ATLA because of the presence of cytoplasmic IgG in cells of these lines. Therefore, mouse monoclonal antibodies (GIN-14 and GIN-2) raised against one (p28) of at least six polypeptides of ATLA in MT-2 cells¹³ used in combination with fluorescein isothiocyanate (FITC)-conjugated goat anti-mouse IgG. These monoclonal antibodies reacted with ATLA-positive but not -negative cell lines. (Y.T. *et al.*, in preparation.) All seven ATL-derived cell lines gave a positive reaction in both EBNA and ATLA staining with three reference sera from anti-EBV reactive healthy persons and with two monoclonal antibodies to ATLA, respectively. They failed to react with three negative control sera or with supernatants from control cultures. From 0.1 to 7% cells were consistently ATLA positive, whereas more than 90% of cells showed EBNA staining. The sera did not react with cells of 17 ATL-unrelated lines of haematopoietic origin⁴.

These data demonstrate that all lines established from the seven patients with ATL express ATLA in a small fraction of the cells and EBNA in the majority of cells. This could indicate that ATLA was expressed in a small number of T cells, although most of the cells in these lines were negative for E rosettes, whereas EBNA was present in a much larger number of B cells. To clarify this, cells of two lines, ATL 1 and 2, were cloned by using Seaplaque agarose (Marine Colloids) as described previously¹⁴. Three weeks after inoculation colonies of varying size were obtained in soft agarose. Cloning efficiencies ranged from 0.1% and 3.1% in ATL 1 and 2 lines, respectively. Clones of ATL 2 cells were isolated by Pasteur pipette and subcultured for 7 days in liquid medium. Then, they were stained for ATLA and EBNA. Table 3 clearly shows that two-thirds of the clones are positive for ATLA and they are all positive for EBNA at the same time. Most of them had sIg (IgG + IgM) ranging from 0.1 to 60% of the total cell population (data not shown). Analogous results were obtained by analysing clones of ATL 1 cells. These data show that ATLA and EBNA are not mutually exclusive and present in the same B-cell population expressing sIg.

It is interesting to analyse the infection of B lymphoblasts with ATL and EBV. It is well established that EBV is present in B lymphocytes *in vivo* and can infect them *in vitro*¹. However infectivity of ATL seems not to be restricted to T cells in view of the following preliminary observations: (1) Not only T but also B, non-T and non-B cells either of primary or of established cell lines were infectable with ATL by cocultivation with an ATL-producer T-cell line, MT-2 (ref. 9 and N.Y. unpublished observation). (2) Proviral DNA of ATL was found after cell fractionation in B-cell-enriched fractions as well as in T-cell fractions of the blood from ATL patients (M. Yoshida, personal communication).

Table 1 Clinical and laboratory data of patients with ATL

Case no.	Specimen	Sample date	Age/sex	White blood count	% Leukaemic cells	% Cells with		Antibodies against	
						E-RFC	sIg	EBV-VCA	ATLA
1	PB	9/80	54/F	66,400	80	94.6	1.2	160	160
2	PB	1/81	39/F	148,600	83	93.1	0.7	ND	ND
3	PB	5/81	65/F	23,300	40	95.6	2.8	80	10
4	LN	2/81	63/F	6,800	1	74.9 (84.6)	2.7 (0.9)	160	160
5	PB	12/81	55/M	35,700	62	43.0	0.4	ND	NS
6	LN	12/81	66/F	121,600	87	76.0 (14.2)	6.3 (0.4)	160	80
7	PB	12/81	67/F	12,000	75	85.8	2.0	80	40

PB, peripheral blood; LN, lymph node; ND, not done; NS, nonspecific fluorescence. Number in parentheses shows a result of lymph node.

Table 2 Characteristics of ATL cell lines from ATL patients

Cell line	Months after establishment	% Cells with			% Of fluorescent cells	
		E-R	sIgG	sIgM	ATLA*	EBNA
ATLB 1	20	<1	<1	55	1~2	>90
ATLB 2	15	<1	35	5	2~4	>90
ATLB 3	10	<1	18	40	3~4	>90
ATLB 4	13	<1	<1	67	5~7	>90
ATLB 5	5	<1	<1	5	0.1~0.2	>90
ATLB 6	5	<1	ND	ND	0.1~0.2	>90
ATLB 7	5	<1	<1	75	<0.1	>90

E-R, receptor for sheep erythrocytes.

* ATLA-positive cells were determined by mouse monoclonal antibodies (GIN-14 and GIN-2) as described in the text. The use of both antibodies resulted in very similar percentage of immunofluorescence-positive cells.

Available data indicate that it is hard to obtain continuous cultures of ATL-positive T-cell lines from patients with ATL. On the other hand, it is well established that EBNA-positive B-cell lines are easily obtained from patients with African Burkitt's lymphoma and infectious mononucleosis or even from normal EBV-seropositive persons by the activation of residual EBV genomes¹. These data suggest that ATL-positive T cells do not have the growth capacity in the present conditions of *in vitro* cultivation while EBV-positive B cells grow easily in tissue culture. It is likely that proliferation of ATL-positive leukaemic T cells *in vivo* are fully sustained by factors present in the serum whereas EBNA-positive B cells are strictly suppressed in ATL patients as in normal EBV-seropositive donors.

Van der Loo *et al.*³ and later Gallo and his colleagues independently found the C-type virus from patients with T-cell malignancies, mycosis fungoides and Sézary's syndrome⁴⁻⁶. The latter group named their virus isolates human T-cell leukaemia

virus (HTLV) and characterized them extensively. Importantly, HTLV was reactive with serum from Japanese ATL patients. They also showed that Japanese ATL and Western ATL were clinically and haematologically indistinguishable, and that the retrovirus of Japanese ATL was also indistinguishable from Western isolates in serological studies¹⁵⁻¹⁷. These data indicate that HTLV is closely related, if not identical, with ATL. A worldwide survey of the distribution of the virus is required.

The monoclonal antibody (GIN-14) used in the present studies only reacted with ATLA-positive but not -negative cell lines. This mouse antibody precipitated at least the p28 polypeptide of the ATLA complex from ATL-positive cells. The p28 polypeptide was shown to be precipitated consistently from ³⁵S-methionine-labelled ATL-positive cells when anti-ATLA-positive human sera were reacted¹³. Moreover, another monoclonal antibody (GIN-2) precipitated both the p28 and p19 polypeptides at the same time. This suggests that p28 and p19 polypeptides are related antigenically. Tryptic peptide analysis of these polypeptides are in progress. Preliminary data by the use of metabolic labelling also showed that one of the ATLB 2 clones synthesized p24 which is a major protein of ATL^{7,13} by the use of human anti-ATLA-positive sera.

Our data may raise the suspicion that B cells act as the ATL reservoir in infected individuals. ATL may initially infect any class of lymphocyte or stem cell of haematopoietic origin before or after the EBV infection but an interaction between type C and herpes viruses occurs as suggested by previous observations^{1,18,19}. So far ATL patients without antibodies to EBV have not been detected in our laboratory where sera from more than 100 cases have been examined. Further studies will clarify whether or not the interaction of EBV and ATL is important in the etiology of ATL.

We thank H. zur Hausen for critical review of the manuscript. This work was supported by grants-in-aid from the Ministry of Education, Science and Culture, and from the Ministry of Health and Welfare of Japan.

Table 3 Per cent of ATLA- and EBNA-positive cells in cloned ATLB 2 cells

Clone no.	% Cells with		Clone no.	% Cells with	
	ATLA	EBNA		ATLA	EBNA
1	35	>90	16	<1	>90
2	<1	>90	17	35	>90
3	0	>90	18	3	>90
4	5	>90	19	0	>90
5	0	>90	20	<1	>90
6	<1	>90	21	40	>90
7	40	>90	22	<1	>90
8	<1	>90	23	40	>90
9	2	>90	24	<1	>90
10	4	>90	25	0	>90
11	0	>90	26	0	>90
12	0	>90	27	8	>90
13	8	>90	28	40	>90
14	0	>90	29	0	>90
15	<1	>90	30	13	>90

Received 25 May; accepted 4 August 1982.

- Epstein, M. A. & Achong, B. G. (eds) *The Epstein-Barr Virus* (Springer, Berlin, 1979).
- zur Hausen, H. & Gissmann, L. *Viral Oncology* (ed. Klein, G.) 433-447 (Raven, New York, 1980).
- Van der Loo, E. M. et al. *Virchows Arch. B Cell Path.* **31**, 193-203 (1979).
- Poiesz, B. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7415-7419 (1980).
- Reitz, M. S., Poiesz, B. J., Ruscetti, F. W. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1887-1891 (1981).
- Poiesz, B. J., Ruscetti, F. W., Reitz, M. S., Kalyanaraman, V. S. & Gallo, R. C. *Nature* **294**, 268-271 (1981).
- Hinuma, Y. et al. *Proc. natn. Acad. Sci. U.S.A.* **78**, 6476-6480 (1981).
- Yoshida, M., Miyoshi, I. & Hinuma, Y. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2031-2035 (1982).
- Yamamoto, N., Okada, M., Koyanagi, Y., Kannagi, M. & Hinuma, Y. *Science* **217**, 737-739 (1982).
- Reedman, B. M. & Klein, G. *Int. J. Cancer* **11**, 499-520 (1973).
- The T- and B-Cell Malignancy Study Group Japan. *J. clin. Oncol.* **11**, 11 (1981).
- Morgan, D. A., Ruscetti, F. W. & Gallo, R. C. *Science* **193**, 1007-1008 (1976).
- Yamamoto, N. & Hinuma, Y. *Int. J. Cancer* (in the press).
- Yamamoto, N. & zur Hausen, H. *Nature* **280**, 244-245 (1979).
- Catovsky, D. et al. *Lancet* **i**, 639-643 (1982).
- Robert-Guroff, M. et al. *Science* **215**, 975-978 (1982).
- Gallo, R. C. et al. *Lancet* **i**, 683 (1982).
- Kufe, D. et al. *Proc. natn. Acad. Sci. U.S.A.* **70**, 737-741 (1973).
- Markham, P. D., Ruscetti, F. W., Salahuddin, S. Z., Gallagher, R. E. & Gallo, R. C. *Int. J. Cancer* **23**, 148-156 (1979).

T4 late transcripts are initiated near a conserved DNA sequence

Alan C. Christensen & Elton T. Young

Department of Biochemistry SJ-70, University of Washington, Seattle, Washington 98195, USA

Bacteriophage T4 late transcription is unusual, among prokaryotes, in its complexity. Late transcription requires the host RNA polymerase, the products of T4 genes 33, 45 and 55, and other small polypeptides, the genes of which have not been identified¹. In addition the DNA template must be 'competent' for late transcription. First the DNA must contain the substituted base 5-hydroxymethyl cytosine in place of cytosine (this requirement is eliminated by a mutation in the T4 *alc* gene)^{1,2}. Second, the DNA must be replicating, although late transcription can be uncoupled from DNA replication by mutations in the T4 genes coding for DNA ligase (gene 30) and DNA exonuclease (gene 46)^{1,3}. We report here the location of the initiation sites of the messenger RNAs (mRNAs) synthesized *in vivo* from four late genes (genes 21, 22, 23 and 36) by S₁ nuclease mapping and we have determined the DNA sequences at these sites. We have found strong homology to the sequence TATAAATAC-TATT immediately upstream from the 5' ends of the late messages and we suggest that this sequence is specifically recognized by the complex responsible for late transcription. Also, we have examined gene 23 mRNA synthesized in the absence of DNA replication using the 30⁻46⁻ mutant described above and find that it is identical to the true late transcript synthesized in normal infections.

Previous work has suggested that gene 23 (which codes for the major capsid protein) is transcribed into a monocistronic message and into two or more polycistronic messages⁴. Transcription has been shown to be initiated upstream from genes 21, 22 and 23, with all these transcripts extending into gene 23 (refs 4, 5). Much of the DNA of this region has been sequenced (our unpublished data, ref. 6; M. Parker, personal communication; M. Showe, personal communication). It has also been suggested that genes 36, 37 and 38 are cotranscribed⁷ and part of this group of genes has also recently been sequenced⁸. These regions of the T4 genome are shown in Fig. 1, with locations of the genes indicated. The restriction fragments used as probes for S₁ nuclease mapping^{9,10} and DNA sequencing are indicated, and are labelled 1 to 4. These fragments were isolated, strand separated, 5' end-labelled, and subjected to the Maxam and Gilbert¹¹ sequencing reactions.

Portions of the same labelled DNA preparation were hybridized to RNA extracted 20 min after infection from T4-infected *Escherichia coli*. The DNA-RNA hybrids were then treated with S₁ nuclease to digest all of the DNA that was not in a

hybrid form with an RNA molecule. The DNAs were then resolved on a DNA sequencing gel and the gel was subjected to autoradiography (Fig. 2). Lanes G, A, T and C represent the results of Maxam and Gilbert sequencing reactions. DNA in the O lanes was incubated without S₁. In all cases there is a low level of background at every position in these lanes with the majority of the radioactivity remaining in the full length fragment. DNA in the N lanes was incubated in the absence of RNA and then digested with S₁. There was no radioactivity evident in these lanes, except for a variable amount of completely undigested probe remaining at the top of the lane due to the incomplete digestion by S₁. DNA in the R lanes was hybridized with late RNA and treated with S₁. These lanes have bands corresponding to 5' ends of RNA molecules. The bands are of an intensity that correlates with the amount of RNA added (data not shown). There is generally a cluster of bands in the R lanes which is probably due to variable cleavage at the junction between single-stranded DNA and DNA-RNA hybrid. It is, therefore, impossible to determine exactly which nucleotide corresponds to the first nucleotide of the mRNA, but it is probably within two or three nucleotides of the major band of each cluster. A correction must also be made because the sequencing reactions remove the base in question and leave a 3'-phosphate, while S₁ leaves the base intact and creates 3'-hydroxyl ends¹². The sense strand sequences surrounding the 5' ends of the late transcripts are shown in Fig. 3. The sense strand was independently sequenced in every case. The lengths of the tails of the arrows in the figure are proportional to the intensity of the corresponding band as measured by densitometry.

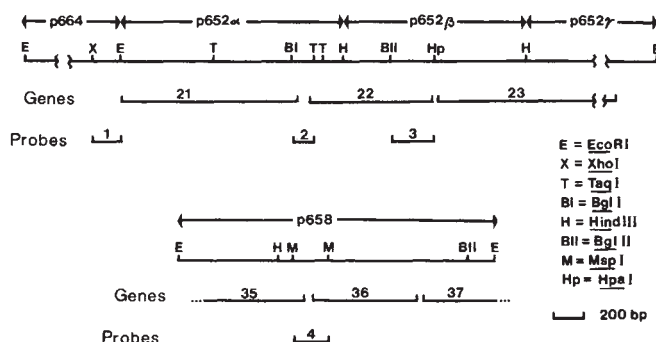
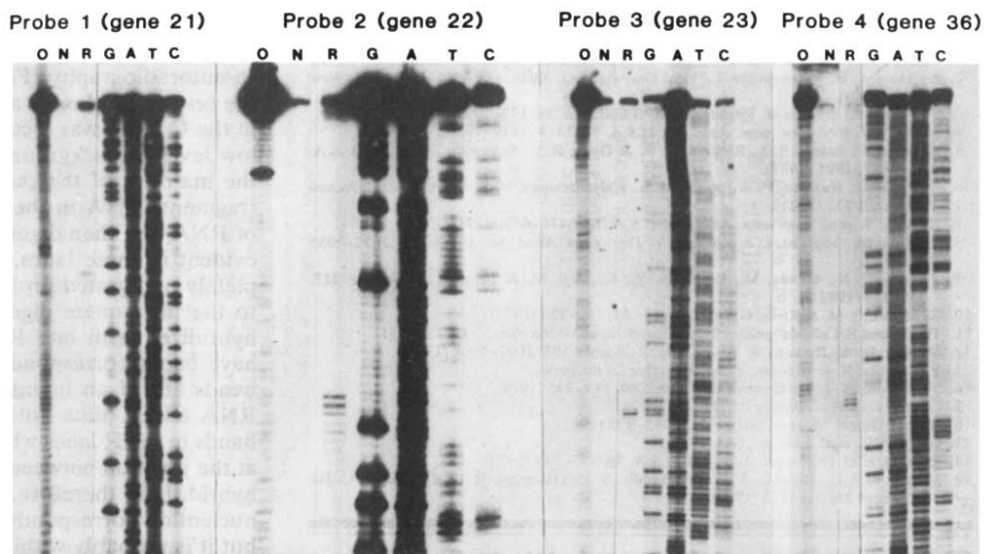


Fig. 1 Restriction maps. Clones were isolated as described in Mattson *et al.*²¹ and Young *et al.*²². The *EcoRI* fragment containing genes 18, 19, 20, PIP (ref. 23) and part of 21 is called 664. The *EcoRI* fragment containing genes 21 to 23 is 652. This fragment was subcloned with *HindIII* into pBR313, forming 652A, 652B and 652C as previously described^{4,22}. The T4 inserts of 652A, 652B and 652C were transferred into pBR322 and renamed 652 α , 652 β and 652 γ . The *EcoRI* fragment containing genes 35 to 37 is 658. Genes 22 and 23 are placed by DNA sequence data (for gene 22, M. Parker, personal communication and our unpublished data; for gene 23, ref. 6). Genes 35, 36 and 37 are placed by comparison to published DNA sequence data⁸. The location of gene 21 is estimated from genetic data and the size of the polypeptide (M. Showe, personal communication and ref. 24). Fragments used for sequencing and S₁ mapping were isolated as described below. Enzymes were obtained from Bethesda Research Labs or New England Biolabs. Agarose was from SeaKem; acrylamide and bisacrylamide were from Bio-Rad. Probe 1: Plasmid 664 was digested with *XhoI* and *EcoRI* and the 170-base pair (bp) band purified by electrophoresis in a 5% polyacrylamide gel (20:1 acrylamide: bisacrylamide) in TBE buffer (90 mM Tris, 90 mM Boric Acid, 1 mM EDTA) followed by electroelution in $\frac{1}{2} \times$ TBE buffer at 50 V overnight. The electroeluate was loaded onto a Whatman DE-52 column (1 cm diameter, bed volume 1 ml) in 50 mM Tris, 10 mM EDTA, 150 mM NaCl, pH 7.4. The DNA was eluted from the column with 50 mM Tris, 10 mM EDTA, 1M NaCl, pH 7.4, ethanol precipitated, washed with 80% ethanol, dried in vacuum and resuspended in 10 mM Tris, 1 mM EDTA, pH 8.0. Probe 2: Plasmid 652 α was digested with *BglI* and *TaqI* and the 140-bp fragment was isolated from a 5 per cent polyacrylamide gel as described above. Probe 3: Plasmid 652 β was digested with *HpaI* and *BglII* and the 275-bp fragment was isolated from a 5 per cent polyacrylamide gel as described above. Probe 4: Plasmid 658 was digested with *HindIII* and *BglII*. The 1,200-bp fragment was resolved on a 1% agarose gel and electroeluted as above. The purified fragment was then digested with *MspI* and the 220-bp fragment was isolated from a 5% polyacrylamide gel as described above.

Fig. 2 Maxam–Gilbert sequencing and S_1 nuclease mapping. Probes were obtained as described in Fig. 1 legend. DNA fragments were strand separated on a 5% polyacrylamide gel as described in Maxam and Gilbert¹¹. Single-stranded DNA fragments were identified by staining the gel in 0.5 $\mu\text{g ml}^{-1}$ ethidium bromide, and the gel slices were cut out and electroeluted as described in Fig. 1 legend. The electroeluate was precipitated with isopropanol, washed with 80% ethanol, dried in vacuum, resuspended in 0.3 M sodium acetate and transferred to a 1.5 ml Eppendorf tube. Single-stranded DNA fragments were 5' end-labelled as follows: DNA (about 10–20 pmol of 5' ends) was ethanol precipitated and washed with 80% ethanol. 100 μCi [γ -³²P]ATP (specific activity 1,000–3,000 Ci mmol⁻¹ in aqueous solution, NEN) was added to the wet DNA pellet and the mixture was dried in vacuum. The DNA and ATP were resuspended in water and stock solutions were added to give a final volume of 50 μl and final concentrations of 50 mM Tris pH 7.6, 10 mM MgCl_2 , 100 μM EDTA, 50 mM dithiothreitol, 1 mM spermidine HCl, 300 μM ADP. Final concentrations of DNA 5' ends and [γ -³²P]ATP were approximately 0.2 μM and 1 μM , respectively. Five units of polynucleotide kinase (New England Biolabs) was then added. Incubation was at 37 °C for 1 h, followed by addition of 200 μl 2.5 M ammonium acetate and 750 μl ethanol. The sample was precipitated (at –70 °C) twice, washed with 80% ethanol and dried. The DNA was then resuspended in 10 mM Tris, 1 mM EDTA, pH 8.0 and distributed into tubes for S_1 mapping and DNA sequencing. The G and C reaction tubes received half as much of the DNA sample as any of the other tubes. DNA sequencing reactions were done as in Maxam and Gilbert¹¹ except the G+A reaction. For this reaction the labelled DNA was resuspended in 10 μl H₂O. Then 3 μl of 10% formic acid was added and the tube was incubated at 37 °C for 10 min. To stop the reaction 200 μl of ice cold 0.3 M sodium acetate containing 5 μg transfer RNA was added, followed by 750 μl of ethanol. All samples were precipitated three times, washed with 80% ethanol and dried. Piperidine cleavages were done as described in Maxam and Gilbert¹¹, except that the samples were transferred to fresh tubes following the 90 °C incubation and four drying steps were done with the addition of 40 μl H₂O before the last three. T4 late RNA was prepared by infecting *E. coli* B⁺ at 4×10^8 cells ml⁻¹ with T4D at a multiplicity of infection of 10, at 30 °C. RNA was extracted 20 min after infection by the method of Hagen and Young²⁵. S_1 nuclease mapping was done by the method of Nasmyth *et al.*¹⁰ using 192 U of S_1 nuclease (Bethesda Research Labs) per reaction. Samples were run on 35 cm $\times$ 43 cm $\times$ 0.4 mm gels (Bethesda, sequencing gel apparatus SO) consisting of 7.6 per cent acrylamide, 0.4 per cent bisacrylamide, 50 per cent urea in TBE buffer at 65 W per gel. Probes numbered as in Fig. 1 were used as indicated at the top of the figure. Lanes G, A, T and C are DNA sequencing reactions G, G+A, T+C and C, respectively. Lanes O, N and R are S_1 mapping experiments. Lane O is a no S_1 control using 30 μg of late RNA, lane N is a no RNA control and lane R was a reaction done with 30 μg of late RNA. Autoradiography was on Kodak XAR film.



The sequences are aligned to emphasize the absolutely conserved sequence TATAAATA and the highly conserved CTATT immediately following. The sequences were determined further upstream (to at least –80 with respect to the 5' end of the mRNA) and no significant homology was observed (data not shown). Nor is there any similarity at –35, either to one another or to the –35 sequence of *E. coli* promoters. As far as we are aware, this conserved sequence is not found anywhere in the T4 DNA sequenced to date, with the following exceptions: (1) the four cases mentioned here; (2) in front of the late genes 24 (M. Parker, personal communication; T. Elliott, personal communication), and *e* (lysozyme)¹³; and (3) in front of gene 32 (ref 26).

That these 5' ends are due to transcription initiation rather than mRNA processing has been shown in two of these cases. Kassavetis and Geiduschek⁵ have shown that there are RNA molecules with 5'-diphosphate or triphosphate termini, that have most likely arisen through initiation, whose 5' ends map at essentially the same positions as the 5' ends mapped in Fig. 2, lanes 2R and 3R (upstream of genes 22 and 23, respectively).

We have examined the mRNA produced in an infection with the T4 triple mutant *tsL56* (gene 43, DNA polymerase) *amH39X* (gene 30, DNA ligase) *amN130* (gene 46, DNA exonuclease). When this mutant infects *E. coli* at 30 °C, DNA replication and late transcription both occur. At the restrictive temperature of 43 °C, DNA replication is blocked but late transcription occurs even in the absence of replication because of the mutations in genes 30 and 46 (ref. 3). One hypothesis to explain this uncoupling of late transcription from DNA replication is that nicks are transiently formed in the DNA during replication. The absence of ligase mimicks this aspect of DNA replication by allowing nicks which form in the DNA to remain unsealed. The mutation in DNA exonuclease merely serves to prevent degradation of the infecting DNA³. S_1 nuclease mapping using probe 3 (gene 23) and RNA extracted from T4 *tsL56 amH39X amN130*-infected cells is shown in

Fig. 4. It can be seen that there is no homologous RNA produced at early times at 30 °C (lane W) or at 43 °C (lane Y) and that the RNA synthesized at late times at 43 °C (lane Z) appears identical to the late RNA of the wild-type infection (lane R) or the mutant infection at 30 °C (lane X). This shows that 5' ends of gene 23 late transcripts in the replication uncoupled mutant *tsL56 amH39X amN130* are identical to those of normal replication-dependent transcripts. This means that random nicks in the DNA do not serve as sites for initiation of transcription. Assuming that this is the case for all late genes, two possible conclusions can be drawn: (1) there is an enzyme which nicks the DNA at specific sites. These nicks serve as initiation sites for RNA polymerase¹⁴. DNA ligase seals these nicks, thereby preventing transcription initiation and DNA replication either prevents the ligase-mediated sealing or greatly increases the occurrence of these nicks; (2) either DNA replication, or random nicking of the DNA in the absence of both replication and ligase, alters the topology of the DNA template, allowing initiation of late transcription at specific promoter sequences.

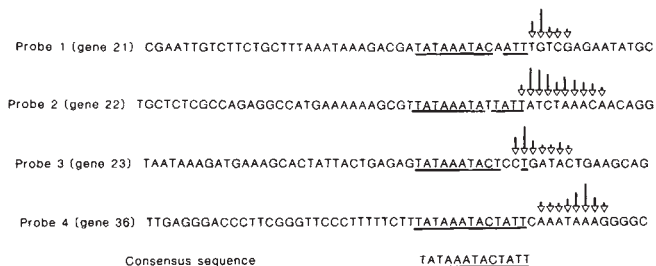


Fig. 3 DNA sequences of the mRNA identical (sense) strand near the positions of 5' ends revealed by S_1 mapping. Densitometry of the bands in lanes R of Fig. 2 was done using a Joyce-Loebl scanning densitometer with integrator. The lengths of the tails of the arrows are proportional to the measured band density. The consensus sequence is indicated at the bottom of the figure and bases identical to it are underlined in the DNA sequences.

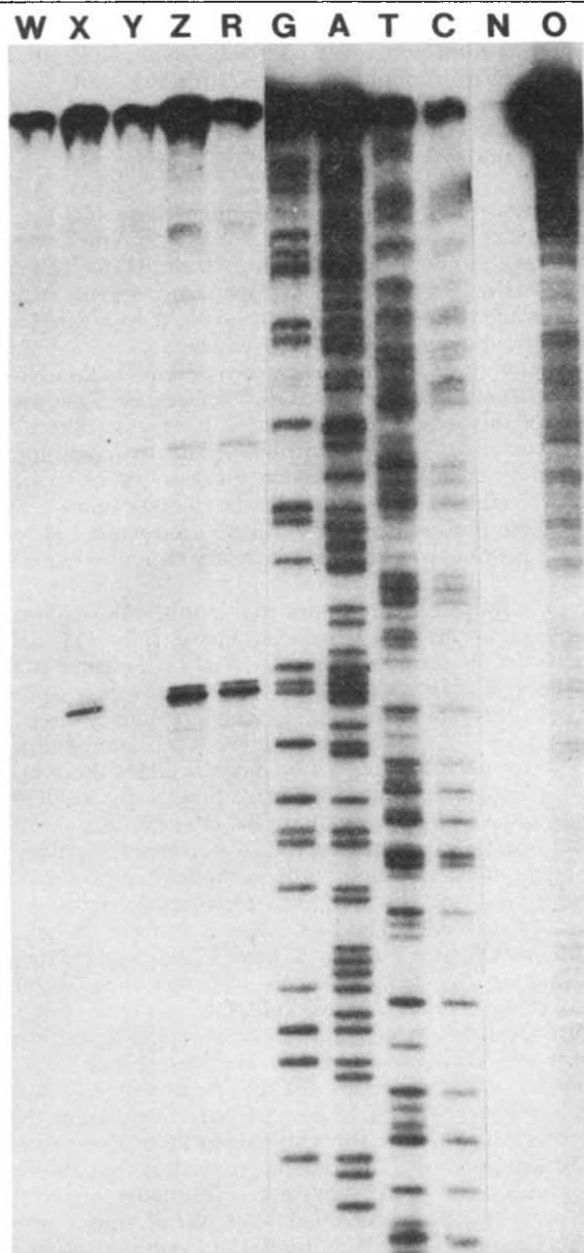


Fig. 4 Replication-uncoupled transcription. Probe 3 was used (gene 23). Lanes G, A, T, C, N, O and R are as described in Fig. 2. Lanes W, X, Y and Z are S_1 mapping experiments performed as described in Fig. 2 using RNA extracted from *E. coli* infected with the T4 mutant *tsL56 amH39X amN130*. Lanes W and X are from an infection of *E. coli* B^F at 30 °C with RNA extracted at 5 and 20 min after infection, respectively. Lanes Y and Z are from an infection of *E. coli* B^E at 43 °C with RNA extracted at 3 and 18 min after infection, respectively.

Transcriptional specificity in prokaryotes is conferred in large part by DNA sequences at -10 (Pribnow box) and -35 relative to the initiation site¹⁵⁻¹⁷. Temporal control of gene expression in the bacteriophage life cycle can be achieved by synthesis of a new RNA polymerase molecule with different specificity. For example, T7 directs the synthesis of an RNA polymerase which does not recognize the -35 and -10 regions of *E. coli* or T7 class I promoters, but instead recognizes a specific 23-base pair sequence¹⁸. Alternatively, phage-specified proteins can modify the existing RNA polymerase, altering its DNA sequence specificity. This phenomenon is well documented in *Bacillus subtilis*. For example, phage SP01 synthesizes gp 28 which displaces the σ subunit of the host RNA polymerase. This activates middle promoters which have -35 and -10 sequences different from those of *B. subtilis* promoters or SP01 early promoters¹⁹. T4 is similar to SP01 in that it modifies the host RNA polymerase rather than synthesizing a new enzyme. Like SP01

middle promoters, the DNA sequence which is well conserved among the T4 late transcripts examined here is different from the sequences recognized by the unmodified RNA polymerase. Although the first five bases of the conserved sequence are identical to the first five bases of the Pribnow box (TATAA), the sixth base of the Pribnow box is a 100 per cent conserved T (refs 15-17) while the 100 per cent conserved A found here is actually more similar to eukaryotic promoter sequences²⁰. The consensus sequence is also longer and closer to the 5' end of the mRNA than the Pribnow box recognized by the host enzyme and there is apparently no specific sequence recognized at -35. The change in transcriptional specificity that occurs at late times is thus not only a substitution of one set of specific DNA-protein interactions with another, but also a fundamental change in the geometry of the interactions.

We thank M. Parker, G. Kassavetis, T. Elliott, E. P. Geiduschek, M. Showe, G. Smith and H. Krisch for communicating results before publication, G. Stormo and L. Gold for access to T4 DNA sequences, M. Parker, A. H. Doermann and C. Furlong for help and technical advice, and the NIH for support (grant AI-09456).

Received 28 April; accepted 13 July 1982.

1. Rabussay, D. & Geiduschek, E. P. in *Comprehensive Virology* Vol. 8 (eds Fraenkel-Conrat, H. & Wagner, R. R.) 1-196 (Plenum, New York, 1977).
2. Snyder, L., Gold, L. & Kutter, E. M. *Proc. natn. Acad. Sci. USA* **69**, 603-607 (1972).
3. Wu, R., Geiduschek, E. P. & Cascino, A. J. *molec. Biol.* **96**, 539-562 (1975).
4. Young, E. T., Menard, R. C. & Harada, J. J. *Virology* **40**, 790-799 (1981).
5. Kassavetis, G. A. & Geiduschek, E. P. *EMBO J.* **1**, 107-114 (1982).
6. Parker, M., Christensen, A. C., Young, E. T. & Doermann, A. H. (in preparation).
7. King, J. & Laemmli, U. K. *J. molec. Biol.* **62**, 465-477 (1971).
8. Oliver, D. B. & Crowther, R. A. *J. molec. Biol.* **153**, 545-568 (1981).
9. Berk, A. J. & Sharp, P. A. *Cell* **12**, 721-732 (1977).
10. Nasmyth, K. A., Tatchell, K., Hall, B. D., Astell, C. & Smith, M. *Cold Spring Harb. Symp. quant. Biol.* **45**, 961-981 (1980).
11. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 499-560 (1980).
12. Sollner-Webb, B. & Reeder, R. H. *Cell* **18**, 485-499 (1979).
13. Emrich-Owen, J., Schultz, D. W., Taylor, A. & Smith, G. R. *J. molec. Biol.* (submitted).
14. Lewis, M. K. & Burgess, R. R. *J. biol. Chem.* **255**, 4928-4936 (1980).
15. Pribnow, D. in *Biological Regulation and Development* Vol. 1 (ed. Goldberger, R. F.) 219-278 (Plenum, New York, 1979).
16. Rosenberg, M. & Court, D. A. *Rev. Genet.* **13**, 319-353 (1979).
17. Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269-281 (1981).
18. Rosa, M. D. *Cell* **16**, 815-825 (1979).
19. Lee, G. & Pero, J. *J. molec. Biol.* **152**, 247-265 (1981).
20. Grosveld, G. C., deBoer, E., Shewmaker, C. K. & Flavell, R. A. *Nature* **295**, 120-126 (1982).
21. Mattson, T., Van Houwe, G., Bolle, A., Selzer, G. & Epstein, R. *Molec. gen. Genet.* **154**, 319-326 (1977).
22. Young, E. T. *et al. J. molec. Biol.* **183**, 423-445 (1980).
23. Völcker, T. A. & Showe, M. K. *Molec. gen. Genet.* **177**, 447-452 (1980).
24. Showe, M. K., Isobe, E. & Onorato, L. J. *J. molec. Biol.* **107**, 35-54 (1976).
25. Hagen, F. & Young, E. T. *J. Virology* **26**, 793-804 (1978).
26. Krisch, H. *et al. Proc. natn. Acad. Sci. USA* (in the press).

The helical hydrophobic moment: a measure of the amphiphilicity of a helix

David Eisenberg, Robert M. Weiss & Thomas C. Terwilliger

Molecular Biology Institute and Department of Chemistry, University of California at Los Angeles, Los Angeles, California 90024, USA

The spatial distribution of the hydrophobic side chains in globular proteins is of considerable interest. It was recognized previously¹ that most of the α -helices of myoglobin and haemoglobin are amphiphilic; that is, one surface of each helix projects mainly hydrophilic side chains, while the opposite surface projects mainly hydrophobic side chains. To quantify the amphiphilicity of a helix, here we define the mean helical hydrophobic moment, $\langle \mu_H \rangle = |\sum_{i=1}^N \bar{H}_i|/N$, to be the mean vector sum of the hydrophobicities $\bar{H}_i$ of the side chains of a helix of N residues. The length of a vector $\bar{H}_i$ is the signed numerical hydrophobicity associated with the type of side chain, and its direction is determined by the orientation of the side chain about the helix axis. A large value of $\langle \mu_H \rangle$ means that the

helix is amphiphilic perpendicular to its axis. We have classified α -helices by plotting their mean helical moment versus the mean hydrophobicity of their residues, and report that transmembrane helices, helices from globular proteins and helices which are believed to seek surfaces between aqueous and non-polar phases, cluster in different regions of such a plot. We suggest that this classification may be useful in identifying helical regions of proteins which bind to the surface of biological membranes. The concept of the hydrophobic moment can be generalized also to non-helical protein structures.

Schiffer and Edmundson² represented helix amphiphilicity by a two-dimensional 'helical wheel' diagram: a projection down the idealized helix axis shows side chains protruding from a circle every 100°. Non-polar residues were mainly on one side of the circle, polar and charged residues on the other. Helical wheels have since been used to represent other amphiphilic helices: for example, it has been noted that portions of apolipoprotein chains³⁻⁵ and a synthetic melittin-like peptide⁶ can be built into helices having one polar face and one non-polar face. The notion of amphiphilic helices has also been helpful in studying the folding of proteins (for example, see ref. 7).

The concept of amphiphilic helices can be quantified by combining a hydrophobicity scale with the helical wheel. Figure 1 shows portions of two helices viewed in projection down their axes. The hydrophobicity of each amino acid residue is represented by a vector directed radially from the projected helix centre to the idealized projected α -carbon position. The length of each vector is H_i , the hydrophobicity of the residue; as this is a signed quantity, hydrophobic residues on one face of the helix reinforce contributions of hydrophilic residues on the other face. For the calculations described in Figs 1 and 2, the values of H_i are those proposed by Janin⁸ on the basis of

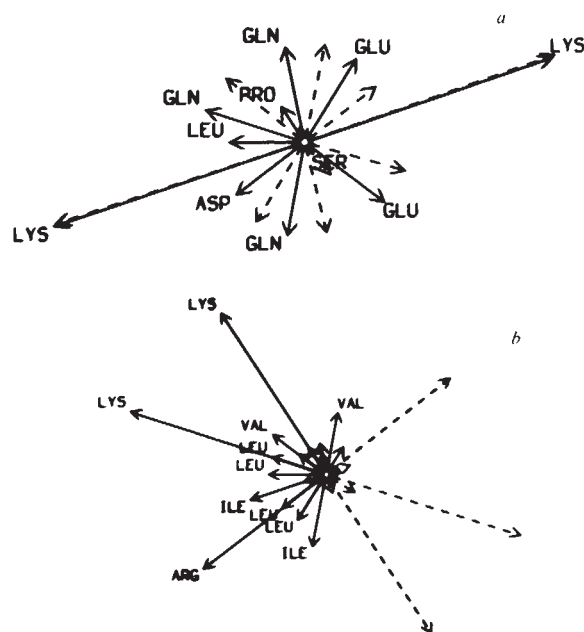


Fig. 1 Graphical representation of the contributions, by residue, to the helical hydrophobic moment, $\bar{\mu}_H$. *a*, The 11-residue stretch of a helix in lactate dehydrogenase starting at residue 308. Hydrophobic residues have positive values for the hydrophobicity, H_i , and are represented as vectors extending out from the centre of the axial projection of the α -helix. Hydrophilic residues have negative values of H_i and their positions about the helix axis are represented by dashed vectors extending from the centre. Their vector contributions are represented by solid lines 180° away. The vector sum of the $\bar{H}_i$ is $\bar{\mu}_H$. As the distribution of solid lines is relatively symmetric, this segment has a small hydrophobic moment, $\langle\mu_H\rangle=0.10$. *b*, Residues 5-22 of melittin, the most amphiphilic 18-residue segment ($\langle\mu_H\rangle=0.40$). Residues of small contribution are represented by vectors but are not labelled. The resulting hydrophobic moment would be off the page to the upper left.

observed distributions of each amino acid between the surface and interior of proteins. We also performed calculations with other sets of hydrophobicity values, including those proposed by Wolfenden *et al.*⁹, von Heijne and Blomberg¹⁰ and Chothia¹¹, as well as with our own set. The general features of the results described below do not depend on the hydrophobicity scale used.

We define the degree of amphiphilic character of a helix perpendicular to its axis by the helical hydrophobic moment, μ_H , the magnitude of the vector sum of the $\bar{H}_i$ for N residues of an α -helix: $|\bar{\mu}_H| = |\sum_{i=1}^N \bar{H}_i|$. To compare moments of helices of different lengths, it is convenient to work with the intensive form of the moment, $\langle\mu_H\rangle = |\bar{\mu}_H|/N$, which we will sometimes refer to simply as the hydrophobic moment. In much of our work, a stretch of 18 residues was chosen because an ideal α -helix of this length makes exactly five turns; thus the side chains are distributed uniformly about its circumference. However other lengths were also studied. Although the real vectors calculated from atomic coordinates (see below) can also be used, the present approach can be used when the tertiary structure is unknown, but the secondary structure can be predicted.

Figure 2 shows a hydrophobic moment plot in which $\langle\mu_H\rangle$ is plotted against the mean hydrophobicity, $\langle H \rangle = (\sum_{i=1}^N H_i)/N$, for helices of N residues, where N is always greater than 10. For helices of ≥ 19 residues, the largest value of $\langle\mu_H\rangle$ for any 18-residue segment is also plotted with a related symbol. Thus, for each helix of 11-18 residues, there is a single point on the plot, and for each helix of 19 or more residues there are two points. The abscissa value of Fig. 2 reflects the solubility of each helix in a non-polar medium, the points falling to the right representing helices which prefer a non-polar medium to a polar medium. The ordinate reflects the tendency of a helix to assume a preferred orientation at an interface between polar and non-polar media. The plot shows data from 64 helices in 26 proteins. Of these helices, 28 have 19 or more residues (up to a maximum of 53 residues for the haemagglutinin membrane glycoprotein of influenza virus) and 36 have 11-18 residues.

Figure 2 shows that different types of helices cluster in different regions of the diagram. Transmembrane sequences are found in the lower right; this is to be expected because, as noted by others¹², such sequences have high mean hydrophobicities which render them soluble in lipid. They also have small helical hydrophobic moments, indicating that all orientations about the helix axis are essentially energetically equivalent. We have assumed that these transmembrane sequences are in α -helical conformations on the basis of Henderson's argument¹³ that this conformation is strongly favoured in a membrane. The common length of 18-24 hydrophobic residues also suggests that these sequences bridge the apolar portion of the bilayer (~ 30 Å) in α -helical conformations (1.5 Å per residue along the helix axis).

In the left and lower central regions of Fig. 2 are clustered α -helical segments from globular proteins. These segments (Table 1) include every representative α -helix of known conformation having 18 or more residues, and every helix of 11-17 residues in the same proteins, after Feldman¹⁴. Of course, other regions of these polypeptide chains are generally not α -helical (and therefore are not plotted). Because these helices are from soluble proteins, it is not surprising that their mean hydrophobicities are smaller than those of membrane-penetrating sequences (that is, they are more hydrophilic). The helical hydrophobic moments for these helices vary substantially, but are often small.

In the upper right of Fig. 2 is a group of helices (represented by triangles and arrowheads) which have both high helical moments and large mean hydrophobicities. The main members of this group are 26-residue peptides known to be both lytic and surface-active. Two are melittins, the main protein components of venom from bees^{15,16}, another is a synthetic melittin-like peptide⁶ and two others are δ -haemolysins, lytic peptides secreted by *Staphylococcus aureus* which are similar to melittin

Table 1 Helices included in the hydrophobic moment plot of Fig. 2

Protein	Organism/organ	1st residue	No. of residues	Evidence for helix*	Protein type	Ref. for sequence or structure
Adenylate kinase	Pig	69(Leu)	15	X ray	Globular	14
		123(Glu)	11			
		144(Glu)	21			
		179(Val)	16			
Alcohol dehydrogenase	Horse	170(Cys)	18	X ray	Globular	14
		202(Gly)	11			
		324(Ser)	13			
		353(Glu)	13			
Carboxy-peptidase A	Cow	14(Thr)	15	X ray	Globular	14
		112(Asn)	11			
		173(Glu)	15			
		215(Asp)	17			
		285(Gln)	22			
Chymotrypsin	Cow	230(Arg)	16	X ray	Globular	14
HA2 haemagglutinin	Influenza virus	77(Glu)	53	X ray	Globular	22, 23
Lactate dehydrogenase	Dogfish	33(Val)	12	X ray	Globular	14
		55(Met)	16			
		107(Glu)	13			
		120(Phe)	11			
		165(Cys)	17			
		249(Trp)	15			
		308(Lys)	22			
Myoglobin	Sperm whale	1(Val)	19	X ray	Globular	14
		20(Asp)	16			
		58(Ser)	20			
		100(Pro)	19			
		125(Ala)	24			
Myohaemerythrin	Marine worm	18(Tyr)	21	X ray	Globular	14
		40(Ser)	23			
		69(Glu)	19			
		93(Ala)	18			
Ribonuclease S	Cow	3(Thr)	11	X ray	Globular	14
		24(Asn)	11			
Thermolysin	<i>Bacillus thermo- proteolyticus</i>	67(Asp)	21	X ray	Globular	14
		137(Ile)	14			
		160(Glu)	20			
		235(Gly)	12			
		260(Arg)	15			
		281(Phe)	16			
		301(Gln)	12			
TMV coat protein	Tobacco mosaic virus	20(Pro)	13	X ray	Globular	24, 25
		38(Gln)	11			
		74(Ala)	15			
		114(Val)	21			
Triose phosphate isomerase	Chicken	17(Lys)	15	X ray	Globular	14
		44(Pro)	12			
		105(Ser)	16			
		138(Ile)	17			
		177(Thr)	20			
		213(Thr)	11			
Glycophorin	Human erythrocyte	12(Ile)	23	Length	Membrane	26
Glycoprotein	Vesicular stomatitis virus	51(Ser)	24	Length	Membrane	27
HA2 haemagglutinin	Influenza A/Victoria or A/Aichi	185(Trp)	24	Length	Membrane	28
HA2 haemagglutinin	Influenza A/Japan	527(Val) [†]	24	Length	Membrane	23
IgM	B lymphocyte	569(Asn)	26	Length	Membrane	29
Isomaltase	Small intestine	10(Ile)	22	Length	Membrane	30
M13 coat	M13 virus	20(Tyr)	18	Length	Membrane	31
M13 procoat	M13 virus	-20(Ser)	21	Length	Membrane	31
δ-Haemolysin	<i>S. aureus</i>	1(Met)	26	Analogy to melittin	Surface	17
	<i>S. aureus</i> (canine strain)	1(Met) [‡]	26		Surface	17
Melittin	<i>Apis mellifera</i>	1(Gly)	26	X ray	Surface	15, 19
	<i>Apis florea</i>	1(Gly) [§]	26	Analogy to above CD	Surface	16
Cytotoxic peptide with melittin-like activity	Synthetic	1(Leu)	26		Surface	6
Diphtheria toxin peptide	Diphtheria toxin	7(Leu)	26	Prediction from sequence	Surface	20

* X ray means that an α -helical conformation has been established by X-ray crystallographic studies; CD, evidence for helicity comes from examination of circular dichroism; length, for a transmembrane protein, the length of the hydrophobic sequence is suitable for spanning the hydrophobic portion of a lipid bilayer as an α -helix.

[†] Corresponds to residue 185 in the enumeration of ref. 28.

[‡] The sequence differs from that of the above strain of *S. aureus* in four residues.

[§] The sequence differs from that of *A. mellifera* in five residues.

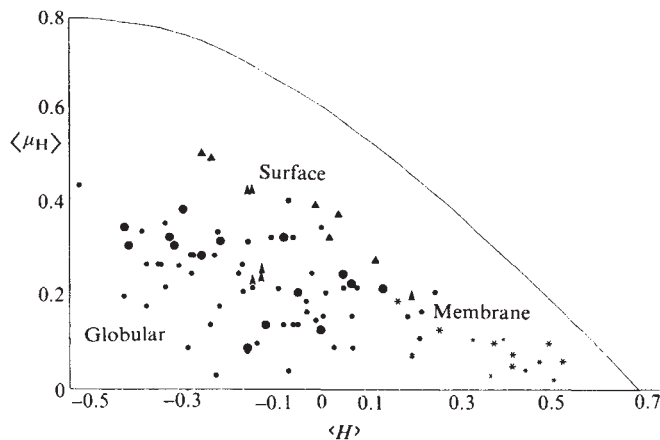


Fig. 2 A hydrophobic moment plot for 64 α -helical segments from the 26 proteins listed in Table 1. The abscissa gives the mean hydrophobicity of each segment and the ordinate gives the corresponding value of $\langle\mu_H\rangle$, as defined in the text. Helices from proteins having different functions plot in various regions of the diagram, as explained in the text. Membrane, membrane-penetrating helices: * represents the average of the entire helix (present for all helices) and x the 18-residue segment having the largest hydrophobic moment (present only for helices of ≥ 19 residues). Globular, helical segments from globular proteins: ●, average and ●, largest moment, Surface, surface-seeking helices: ▲ and ▲. The curve shows the largest possible value of $\langle\mu_H\rangle$ for each value of $\langle H\rangle$, as described in the text.

in both properties and sequence¹⁷. One of the melittins is known from X-ray diffraction studies to have an α -helical conformation^{18,19}; the synthetic melittin is α -helical as judged by circular dichroism⁶. By analogy to these two peptides, we assume here that the δ -haemolysins are also α -helical. The final member of this group is a 26-residue peptide, fragment B from diphtheria toxin. This peptide has been predicted to have an α -helical conformation and is part of a larger CNBr peptide that induces conductance changes in lipid bilayers²⁰. We include it in this group only to suggest that some peptides from larger proteins may have relatively large hydrophobic moments related to special functions.

Also shown in Fig. 2 is the maximum possible helical hydrophobic moment for each value of the hydrophobicity, represented by a curve which meets the abscissa at 0.70, the value for a hypothetical polyisoleucine α -helix. (Cysteine was excluded from the calculations of the maximum hydrophobic moment because the procedure used⁸ to determine hydrophobicities may overestimate that of cysteine.) Note that even the highly amphiphilic δ -haemolysins are far less amphiphilic than some hypothetical helices.

The amphiphilic helical portions of melittin, as well as similar helical regions in other proteins, might be termed 'surface-seeking helices' because their large hydrophobic moments tend to align them at the surface between polar and non-polar phases¹⁹. Another class of proteins believed to function as amphiphilic helices is the apolipoproteins^{3,4}; these have not been included here because even less direct information on their three-dimensional structures is available than for the melittins.

Although proteins of different functional types cluster in different regions of Fig. 2, these regions do not have absolute boundaries. This is expected as the various types of helices do not have entirely distinct functions. For example, the solubility of melittin in aqueous solution is believed¹⁸ to be achieved in part by the highly charged C-terminal segment of the helix. This segment projects polar groups uniformly to all sides¹⁹ and hence lowers the value of $\langle\mu_H\rangle$ for the entire melittin chain below that of the most amphiphilic 18-residue segment. For the two types of melittin and the synthetic melittin-like peptide, this results in arrowheads near the coordinates $\langle H\rangle = -0.15$,

$\langle\mu_H\rangle = 0.25$. Other helical segments that plot in intermediate regions include some of those from the purple membrane protein model described by Engelman *et al.*²¹; these plot between the 'membrane' and 'surface' regions because of the charged and polar residues in the sequence (they are not included in Fig. 2). Such membrane helices with large hydrophobic moments may pair in the membrane, thereby reducing the net hydrophobic moment of the pair. More generally we suspect that segments from a helix that plot in a region of Fig. 2 occupied mainly by another type of helix may have a specialized function.

A potential limitation in the calculation of the helical hydrophobic moment $\langle\mu_H\rangle$ is that amino acid side chains deviate significantly from the idealized positions we have used (spaced 100° apart perpendicular to the helix axis). To clarify this point, we have investigated the properties of the structural hydrophobic moment, $\tilde{\mu}_s$; the most convenient definition for this comparison is $\tilde{\mu}_s = \sum_{i=1}^N H_i \tilde{s}_i$, where $\tilde{s}_i$ is a unit vector pointing from the α -carbon atom of the i th residue to the centre of the residue side chain. This quantity can be calculated for any protein segment whose structure is known; when calculated for the idealized helices used in obtaining the mean hydrophobic moment, the angular deviation between the idealized $\tilde{s}_i$ and the corresponding vectors perpendicular to the helix axis for a real helix averaged 30°. The difference in magnitude between the mean and structural moments, after scaling the latter by the length of the helix, averaged 25%. When computed for a helix, the structural hydrophobic moment $\tilde{\mu}_s$, like μ_H , emphasizes the amphiphilicity perpendicular to the helix axis. An alternative definition is $\tilde{\mu}_s = \sum_{i=1}^N (H_i \tilde{R}_i - \langle H\rangle \tilde{R}_i)$, in which $\tilde{R}_i$ is a vector from any origin to the centre of the side chain of the i th residue. This definition of the hydrophobic moment represents amphiphilicity in directions both parallel and perpendicular to a helix axis and may have applications in energetics calculations.

We conclude that hydrophobic moment plots can reveal correlations between amino acid sequence and protein function; in particular, they can identify sequences particularly well suited for forming surface-seeking helices.

We thank L. Aha for discussions, and acknowledge support from the NSF (PCM 80-03725) and the NIH (GM 16925). T.C.T. was the recipient of a NSF graduate fellowship.

Received 25 February; accepted 24 June 1982.

- Perutz, M. F., Kendrew, J. C. & Watson, H. C. *J. molec. Biol.* **13**, 669-678 (1965).
- Schiffer, M. & Edmundson, A. B. *Biophys. J.* **7**, 121-135 (1967).
- Morrisett, J. D., Jackson, R. L. & Gotto, A. M. Jr *Biochim. biophys. Acta* **472**, 93-133 (1977).
- Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M. Jr *FEBS Lett.* **38**, 247-253 (1974).
- Segrest, J. P. & Feldman, R. J. *Biopolymers* **16**, 2053-2065 (1977).
- DeGrado, W. F., Kezdy, F. J. & Kaiser, E. T. *J. Am. chem. Soc.* **103**, 679-681 (1981).
- Ptitsyn, O. B. & Rashin, A. A. *Biophys. Chem.* **3**, 1-20 (1975).
- Janin, J. *Nature* **277**, 491-492 (1979).
- Woffenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. *Biochemistry* **20**, 849-855 (1981).
- von Heijne, G. & Blomberg, C. *Eur. J. Biochem.* **97**, 175-181 (1979).
- Chothia, C. *J. molec. Biol.* **105**, 1-14 (1976).
- Segrest, J. P. & Feldman, R. J. *J. molec. Biol.* **87**, 853-858 (1974).
- Henderson, R. *Soc. gen. Physiol.* **33**, 3-15 (1979).
- Feldman, R. J. *Atlas of Macromolecular Structure on Microfiche* (Tracor Jitco, Rockville, Maryland, 1976).
- Habermann, E. *Science* **177**, 314-322 (1972).
- Kriel, G. *FEBS Lett.* **33**, 241-244 (1973).
- Fitton, J. E., Dell, A. & Shaw, W. V. *FEBS Lett.* **115**, 209-212 (1980).
- Terwilliger, T. C. & Eisenberg, D. *J. Biol. Chem.* **257**, 6016-6022 (1982).
- Terwilliger, T. C., Weissman, L. & Eisenberg, D. *Biophys. J.* **37**, 353-361 (1982).
- Kayser, G. *et al. Biochem. biophys. Res. Commun.* **99**, 358-363 (1981).
- Engelman, D. M., Henderson, R., McLachlan, A. D. & Wallace, B. A. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2023-2027 (1980).
- Wilson, I. A., Wiley, D. C. & Skehel, J. J. *Nature* **289**, 366-373 (1981).
- Porter, A. G. *et al. Nature* **282**, 471-477 (1979).
- Bloemer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. *Nature* **276**, 362-368 (1978).
- Dayhoff, M. O. *Atlas of Protein Sequence and Structure* Vol. 5, D283 (National Biomedical Research Foundation, Washington DC, 1972).
- Furthmayr, H., Galaray, R. E., Tomita, M. & Marchesi, V. T. *Archs Biochem. Biophys.* **185**, 21-29 (1978).
- Rose, J. K., Welch, W. J., Sefton, B. M., Esch, F. S. & Ling, N. C. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3884-3888 (1980).
- Verhoeven, M. *et al. Nature* **286**, 771-776 (1980).
- Rogers, J. *et al. Cell* **20**, 303-312 (1980).
- Frank, G. *et al. FEBS Lett.* **96**, 183-188 (1978).
- Wickner, W. *Science* **210**, 861-868 (1980).

MATTERS ARISING

BVP models
of nerve membrane

HINDMARSH and Rose¹ propose, for use in simulating networks containing small numbers of neurones, a phenomenological model based on their experimental data, and notable for its simple numerical integrability. Inspection of their paper shows that their experimental data do not in fact extend into the region $x < -15$ mV, yet it is just this region that contains the novel feature of their model, a positive gradient for the isocline $\dot{y} = 0$. Others (for example, see ref. 2) have collected data in this region which do not support the positive gradient used by Hindmarsh and Rose for $\dot{y} = 0$ (see Fig. 1). Thus, their slow variable y has less physical meaning than the usual BVP (Bonhoeffer-van der Pol) slow variable³. Furthermore, for constructing models of this kind, there already exists a method^{4,5} which gives explicitly solvable equations and requires no numerical integration. Over the full range of experimental data, this method gives a better fit than that of Hindmarsh and Rose. The method is as

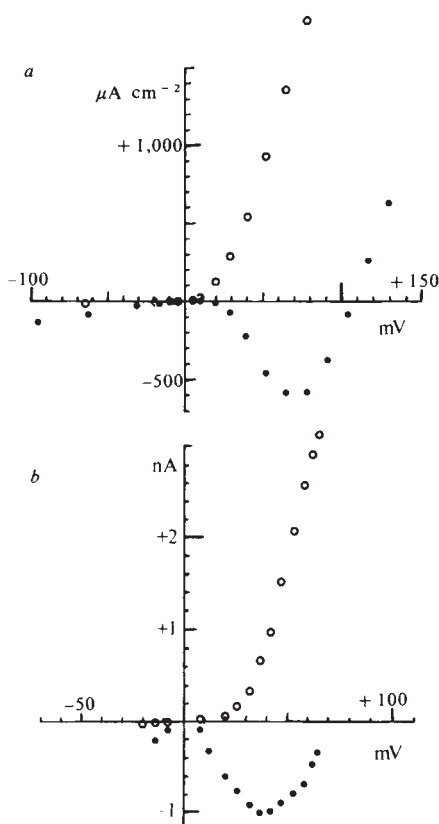


Fig. 1 Schematic summary of the data under consideration. Ordinates show the membrane current supplied by the clamping electrode; the abscissae show the membrane potential at which the clamp is set. $\circ$, Steady-state current; $\bullet$, initial current. *a*, Based on the data of ref. 2 for squid giant axon; *b*, based on the data of ref. 1 for snail neurone soma.

follows: divide the voltage range into three segments; fit the data in each segment with a straight line (making $f(x)$ and $g(x)$ continuous and step-wise linear); and solve the resulting second-order linear system of ordinary differential equations.

Real nerve cell membranes, the Hodgkin-Huxley system⁶, and the usual BVP models⁴ are made relatively more excitable by an anode break shock. In contrast, its novel feature makes the Hindmarsh-Rose model relatively refractory after such a stimulus.

C. J. A. GAME

Department of Neurology,
Academic Hospital of Leiden,
2333 AA Leiden, The Netherlands

1. Hindmarsh, J. L. & Rose, R. M. *Nature* **296**, 162-164 (1982).
2. Hodgkin, A. L., Huxley, A. F. & Katz, B. *J. Physiol., Lond.* **116**, 424-448 (1952).
3. Game, C. J. A. *Biol. Cybernet.* (in the press).
4. Scott, A. C. *Rev. mod. Phys.* **47**, 487-533 (1975).
5. Perkel, D. H. *Comput. biomed. Res.* **9**, 31-43 (1976).
6. Hodgkin, A. L. & Huxley, A. F. *J. Physiol., Lond.* **117**, 500-544 (1952).

HINDMARSH AND ROSE REPLY—The novel feature we wished to draw attention to was not the positive gradient of the isocline $\dot{y} = 0$ in the region $x < 0$, but rather the close proximity of the isoclines $\dot{x} = 0$ and $\dot{y} = 0$. This proximity is determined by $z_p(\infty)$, which is small according to both the Hodgkin-Huxley data for x_p between 0 and -65 mV and our data for x_p between 0 and -20 mV. On the other hand, the location of these isoclines is determined by $z_p(0)$. Few points were obtained for $z_p(0)$ for $x_p < 0$ because by using the single electrode voltage clamp, the initial responses were obscured by a capacitive current. We therefore followed Fitzhugh¹ in taking a cubic form for $z_p(0)$. Due to this, both isoclines have a positive gradient in the region $x < 0$ which, as both Game and Fitzhugh (personal communication) have pointed out, means that our model is relatively refractory after an anode break shock. Although this positive gradient is thus a disadvantage in our model, it is consistent with the following experiment (tail current reversal) performed numerically with the Hodgkin-Huxley² equations.

If the voltage is held at +50 mV until a steady state is reached, and then stepped back to various negative potentials, it is found that when the potential is sufficiently negative, the tail current changes direction. This property is shown, at least qualitatively, by our model, as can be seen from our phase plane diagram³. Given the choice between explanations of tail current reversal or anode break, we chose the former, because we saw that this would also predict the long interspike interval. This suggests that $z_p(0)$ for $x_p < 0$ should be constructed from such

tail current data rather than from negative voltage steps from zero. Such a determination seems more relevant to the time course of the membrane potential during firing.

Clearly, if $z_p(0)$ is determined from voltage clamp steps from zero and the Hodgkin-Huxley data is used (Fig. 1*a*), then straight lines will provide a better fit than a cubic. In this case both isoclines will still have a small positive gradient in the region $x < 0$, which will mean that the phase state on the limit cycle travels far to the left before returning to the origin between the isoclines, resulting in a large undershoot of the membrane potential. In our case, replacing the cubic by three straight lines does not lead to a significant deterioration in accuracy of prediction of the time course of the membrane potential, and does, as Game points out, offer the advantage of explicit solutions.

J. L. HINDMARSH
R. M. ROSE

Department of Applied
Mathematics and Astronomy and
Department of Physiology,
University College,
Cardiff CF1 1XL, UK

1. Fitzhugh, R. *Biophys. J.* **1**, 445-466 (1961).
2. Hodgkin, A. L. & Huxley, A. F. *J. Physiol., Lond.* **117**, 500-554 (1952).
3. Hindmarsh, J. L. & Rose, R. M. *Nature* **296**, 162-164 (1982).

Trophic structure of a grassland
insect community

FROM a detailed study of insect species in a grassland community, Evans and Murdoch¹ concluded that the ratio of herbivorous insects to entomophagous insects remained fairly constant throughout the growing season. They suggested that this reflected an underlying trophic pattern that persisted despite continuing turnover of species. Cole² rejected this claim by showing that an appearance of constancy can be generated by samples taken from a fixed pool of all available species.

It is true that in the absence of exact knowledge of the community structure through the season it is not possible to assert that species turnover occurs. However, it is equally impossible to assert that species turnover does not occur. Nevertheless, this is precisely the arbitrary assumption made by Cole.

He assumes that all species listed by Evans and Murdoch (131 herbivorous, 80 entomophagous) are present throughout the season, and treating the total number of species observed each fortnight as a random sample (r) from the population of 211 species, he demonstrates that the expected number of herbivores agrees

Table 1 Community structure

Time period	Herbivorous species	Entomophagous species	Total no. of species
1	H ₁ , H ₂	E ₁ , E ₂ , E ₃ , E ₄	6
2	H ₃ , H ₄ , H ₅	E ₅ , E ₆ , ..., E ₁₀	9
3	H ₆ , H ₇ , H ₈ , H ₉	E ₁₁ , E ₁₂ , ..., E ₁₈	12
			27

well with the observed number for each fortnight. However, such an agreement can equally arise if the insect community did in fact possess a stable trophic structure. For example, consider the season divided into (say) three time periods and let the exact community structure be as shown in Table 1. Here, each H_i and E_j is a distinct species. There is a complete and known turnover of species, and for each time period the ratio of the number of herbivorous to entomophagous species is exactly 1/2. An assumption like Cole's, that all species are present at all times, will give $n = \text{total species} = 27$, $n_1 = \text{number of herbivores} = 9$, $r_1 = 6$, $r_2 = 9$, $r_3 = 12$; and, for each time period, the expected and observed number of herbivorous species will agree exactly.

Cole's conclusion that the ratio of herbivores to predators, in the insect community studied by Evans and Murdoch, is maintained at a constant level by no other force than a statistical one is thus possible but not proven. The value of Cole's letter is that it gives a warning that an apparent observed trophic structure may not only be due to the background fact of nature, it can also arise for other reasons.

Obviously, much detailed work on the life history of individual species is required before it is possible to assess adequately the extent of species turnover in a given community. Also, even if energy, biomass or nutrients are undeniably transferred from one trophic level to another, it is perhaps an oversimplification to search for general patterns in terms of the number of species, ignoring abundance, average size and requirements of the members of each species.

K. H. LAKHANI

*Institute of Terrestrial Ecology,
Monks Wood Experimental Station,
Abbots Ripton,
Huntingdon PE17 2LS, UK*

1. Evans, F. C. & Murdoch, W. W. *J. Anim. Ecol.* **37**, 259-273 (1968).
2. Cole, B. J. *Nature* **288**, 76-77 (1980).

Thermoluminescence dating of sand dunes

SINGHVI ET AL.¹ have recently suggested that the thermoluminescence (TL) technique can be used successfully to date aeolian sand dunes. These authors use the TL signal from the 1-8 µm fraction of dune sediments in Rajasthan to suggest dune ages of between 2,000 and 20,000 yr. This procedure implicitly

assumes that the 1-8 µm particles are entirely a primary detrital component, and that secondary additions of fines after dune stabilization are unimportant. This assumption warrants close examination, since significant secondary input could result in TL dates which are too young.

Active desert dunes usually contain <3% silt, whereas stabilized and weathered dunes may contain >30%. Goudie *et al.*² have reported 20-30% fines in stabilized dunes from Rajasthan. Post-depositional addition of fines may reflect aeolian dust input³⁻⁵, introduction by surface wash from higher ground^{6,7}, surficial deposition of biogenic silica⁸ or the effects of *in situ* weathering on sand grains⁹. Dust input is probably the most significant factor in terms of errors in TL dating of desert dune deposits. The dust, deposited on a dune surface, may be rapidly eluviated by rainwater percolating through the dune sand column. Wright and Foss¹⁰ demonstrated experimentally that 8 g of fine silt placed on top of a 33-cm column of medium sand was completely leached by <1 litre of water. In sealed sand columns the eluviated fines are deposited as an illuvial sub-surface horizon which grows vertically with time.

As a result of the efficacy of this process, TL dating of 1-8 µm-sized particles may simply reflect rates of aeolian dust addition rather than the age of a dune sand body in which they occur. Where such dust eluviation and deposition occurs, a TL age gradient up the profile is to be expected. Singhvi *et al.* demonstrate such an age gradient in the dunes they studied, but suggest that it may reflect slow vertical accretion of aeolian sand over a substantial period of time. However, in the absence of independent supportive evidence it would be equally valid to suggest that the dune sand bodies were formed entirely in the late Pleistocene, and that TL dates on the upper parts of the dune profiles investigated are artificially low due to contamination. Geomorphological and palaeoenvironmental evidence also suggest that dune formation in Rajasthan has occurred as a series of discrete events in response to periodic climatic changes in the late Pleistocene and Holocene, rather than by slow vertical accretion^{2,11}.

¹⁴C and TL dates on associated archaeological artefacts are of limited value as a cross-check on the reliability of TL dates of silt particles because of the uncertainty surrounding the time relationship between the artefacts and adjacent or surrounding sediment. The best solution to the problem would be to apply the TL dating technique to the sand grains themselves. Both quartz and alkali feldspar inclusions in pottery have been dated using various grain sizes in the range 0.1-0.5 mm and the methodology could be adapted for application to sand dunes. Indeed, several TL laboratories involved in sediment dating already use grains of

≥0.1 mm (ref. 12). Singhvi *et al.* should examine the TL of sand grains at various levels in the dune profiles to establish whether or not they indicate increasing age with depth.

I thank Ann Wintle for comments and discussion.

K. PYE

*Department of Earth Sciences,
University of Cambridge,
Downing Street,
Cambridge CB2 3EQ, UK*

1. Singhvi, A. K., Sharma, Y. P. & Agrawal, D. P. *Nature* **295**, 313-315 (1982).
2. Goudie, A. S., Allchin, B. & Hegde, K. T. M. *Geogr. J.* **139**, 243-257 (1973).
3. Yaalon, D. H. *Soil Sci.* **116**, 146-155 (1973).
4. Syers, J. K., Jackson, M. L., Berkheiser, V. E., Clayton, R. N. & Wex, R. W. *Soil Sci.* **107**, 421-427 (1969).
5. Goudie, A. S. *J. Arid Envir.* **1**, 291-310 (1978).
6. Bigarella, J. J. *Bol. Para. Geosci.* **33**, 133-167 (1975).
7. Walker, T. R., Waugh, B. & Crone, A. J. *Bull. geol. Soc. Am.* **89**, 19-32 (1978).
8. Wilding, L. P. & Drees, L. R. *Clays Clay Miner.* **22**, 295-306 (1974).
9. Goudie, A. S., Cooke, R. U. Doornkamp, J. C. J. *Arid Envir.* **2**, 105-112 (1979).
10. Wright, W. R. & Foss, J. E. *Proc. Soil. Sci. Soc. Am.* **32**, 446-448 (1968).
11. Verstappen, H. T. Z. *Geomorph. Suppl.* **10**, 104-120 (1970).
12. Wintle, A. G. & Huntley, D. J. *Q. Sci. Rev.* **1**, 31-53 (1982).

SINGHVI REPLIES—I thank Dr Pye for his valued comments. In fact, some of the studies he suggested are already in progress and the results will be presented at the Third International Specialist Seminar on Thermoluminescence Dating (Denmark, July 1982) and will be published elsewhere¹. In the meantime, I hope the following information will suffice:

(1) In the field, the permeability of a stabilized dune is circumscribed by factors such as vegetation, carbonate crust and the original fine grain population of the dune sand. In such an arid environment, the experiments of Wright and Foss² (where almost 15 cm of standing water was constrained to move unidirectionally), do not have much relevance. In fact, at Amarpura the average rainfall is <300 mm yr⁻¹.

(2) The archaeological evidence of a habitation and radiocarbon/TL dates indeed validate fine-grain TL dates as being the dates of accumulation of dune sand. Certainly the archaeological material (pottery, charred bone, and so on) cannot percolate down to 1-2 m depths and therefore the overlying sediment has to be post-archaeological debris.

(3) We agree that a comparison between the coarse-sand fraction and the fine-grain TL dates is useful. The same has been attempted with encouraging results for Langhnaj samples. These analyses will be completed shortly and the results will be presented at the TL seminar in Denmark.

A. K. SINGHVI

*Physical Research Laboratory,
Ahmedabad 380 009, India*

1. Singhvi, A. K., Sharma, Y. P. & Agrawal, D. P. *PACT* **9** (in the press).
2. Wright, W. R. & Foss, J. E. *Soil. Sci. Soc. Am. Proc.* **32**, 446-448 (1968).

BOOK REVIEWS

Traits of empire

C.R. Brand

STANDING in an African flower-pot, a pop-eyed, stick-figured ET — displaying an enviable claim to masculinity — looks out from the cover of *The Biology of Human Conduct*. Browsers will presume that some kind of sociobiology or anthropology is in store, and that the Western models of the sub-title are to encounter yet another challenge from the mysterious East. The only clue that matters are otherwise is provided by the author's position in the Department of Experimental Psychology at the University of Oxford — several of whose members are to be put in, together with H. J. Eysenck, to bat for the West, while Pavlov, Teplov, Nebylitsyn and Strelau (in Warsaw) represent the trait-psychology of the East. Thus, in the hope of a crowd, are British psychometrician-psychologists and Russian psychophysiologicalists made to confront each other.

The match turns out to be something of a friendly. All the contestants certainly accept as ground-rules the importance of such questions as: "What are the major (independent) dimensions of personality?"; "What individual differences in psychophysiological mechanisms and processes underlie these behavioural dimensions?"; and "Are such differences partly genetic in origin?". Nor is there radical disagreement as to methods: indeed, some Western readers will be surprised by the prevalence of twin-studies in the countries of the Eastern bloc. As the humanistic fashion of Western psychology decrees, Mangan frequently steps aside to admit that the agreed questions are simple and that reality is complex; but, though Freud, Jung, Piaget and Wilhelm Reich are occasionally called upon to plug holes, the bulk of the book is concerned with the laboratory investigation of Extraversion (alias strength, or non-reactivity, of the nervous system), Neuroticism (alias dynamism) and similar putative dimensions of personality such as "mobility", "lability" and "irradiation of inhibition". Quotations provide the occasional diversion: "Only idealists can speak of thinking without language" (J. V. Stalin, 1951).

Regrettably, the contest staged by the author is a colossal disappointment. Not only does the reader quickly realize that the result is going to be a messy draw, but he learns that Mangan will not even allow there to be a Man of the Match. The preface itself sets the depressive tone by announcing that the book contains "the detritus" of Mangan's 20-year interest in

The Biology of Human Conduct: East-West Models of Temperament and Personality. By G.L. Mangan. Pp.571. ISBN 0-08-026781-5. (Pergamon: 1982.) £35, \$72.

personality; and Mangan soon makes it clear that his regard for the theorists whose work he considers is so beset with scholarly inhibition that he cannot allow the reader the luxury of having any of their theories set out at length. Even his own favourite theme — the importance of the "mobility" dimension — is forever the subject of allusion rather than development; so he can hardly afford to indulge Gray and Strelau. That psychological scientists will have to refer to this reference-packed work or to dismiss the East altogether is the best that can be said.

Mangan's lack of panache may be of situational rather than personological origin. Although the Russians have lately relaxed Stalin's ban on IQ-testing, Soviet typological theory of personality seems to be in a Lysenkoist mess. The fact is that the firmly materialistic and "biological" work that Mangan reports has been tolerated over the years only on the understanding

that it has nothing to do with important psychological differences between people — which differences are considered in the East to be matters of nurture rather than of nature. As Mangan politely puts it when discussing work at the Kiev Institute, "There is no suggestion, of course, of any relationship between nervous system properties and intellectual endowment...". Thus Soviet differential psychophysiologicalists are left to play with somewhat hypothetical variants of Eysenck's E and N, since these dimensions repeatedly manifest themselves even in laboratories that draw on Soviet sophomores for their subjects. But General Intelligence (*g*) is denied them, and they have no certain method for discovering dimensions of personality just because, if they did have one, they might discover *g* for themselves. Mangan's tolerance of this state of affairs allows him to be the dispassionate, if dispiriting, reporter of years of Soviet endeavour — so perhaps the problem with this book is personological after all. □

C.R. Brand is a Lecturer in the Department of Psychology at the University of Edinburgh.

A mixed bag of andesites

Richard J. Arculus

Andesites: Orogenic Andesites and Related Rocks. Edited by R. S. Thorpe. Pp.724. ISBN 0-471-208034-8. (Wiley: 1982.) £134, £59.50.

ANDESITIC volcanoes are forced to the attention of the general public in sometimes unpredictably violent ways involving loss of livelihood and even life. In developed nations, like Japan and the United States, advanced technology is applied to monitoring and forecasting of dangerous eruptions, but given the lengthy repose cycles typical of these volcanoes a great deal remains to be learnt of their behaviour.

In other ways, the products of andesitic volcanoes are fascinating in that their composition and isotopic characteristics provide clues to the nature of the processes that occur in the upper mantle and crust of the Earth during the collision of lithospheric plates. It is possible that the magmatic activity associated with these plate collisions has been an important

agent in the extraction on a geological timescale of continent-forming material from the mantle.

In *Andesites*, R. S. Thorpe has assembled contributions from over 50 authors reviewing many of the intriguing aspects of this type of magmatism. Subjects covered range through classification of andesitic rock types, petrology and mineralogy, chemical and isotopic features, and description of the distribution and characteristics of active andesitic volcanoes excluding only the Vanuatu, Solomon Islands, Philippine and Soviet Union segments of the Pacific rim. Eruptive styles, evolution of andesitic volcanism through time, major types of andesite-associated mineral deposits and experimentally-based studies of petrogenesis are also included.

Clear diagrams and print make the book visually pleasing, and the reader will emerge from the whole volume with an appreciation of some of the main points of current debate on the sources and

evolutionary history of andesites. Unfortunately, the multi-authored approach results in an incoherent overall presentation of the relevant data. For example, several different hypotheses are advanced by individual authors suggesting that andesites are formed from partial melts of subducted lithosphere, from primary melts of metasomatized and hydrated mantle peridotite, or from derivative melts from precursor basaltic magmas. So there is a fair cross-section of expert views, but no balanced, editorial presentation of the defects and successes of these individual hypotheses. Even the definitions of the qualifying terms applied to different types of andesite — "calcalkaline", for instance — are not consistently employed throughout the text.

Uneven coverage of the important aspects of andesitic volcanoes extends unfortunately into one of the most potentially useful parts of the book, that occupied by descriptions of individual arc segments. There is no consistent presentation of information on basement age and character, evolution of volcanism, volumetric abundance of rock types, average and representative chemical and isotopic analyses, mineralogical data and petrogenetic interpretation. Text length devoted to individual arcs is even-handed but unbalanced when the complexities in the 7,000km length of Andean volcanic and plutonic activity merits the same space as the 700km long Lesser Antilles, even if allowance is made for the stimulating variety of oddities that is compressed into the eastern Caribbean Arc.

Nevertheless, the book does contain a number of useful and provocative contributions. Ewart has produced a comprehensive summary of the regional variation of chemistry and mineralogy present in continental and island arc volcanoes, which is a much-needed comparative data base for further study. (However inclusion of the hinterland of the western United States seems unwise in this exercise.) Also instructive is the comparison between the documentation by Ewart that most Cenozoic andesites contain plagioclase phenocrysts with a median composition of 50–60 mole per cent anorthite, and the drastically more sodic plagioclase (25–35 mole per cent anorthite) reported by Condie in Archean andesites. Given other compositional differences such as elevated Ni, Co and Zr contents in Archean andesites, it would appear that direct analogies drawn between modern and ancient, magmatic and tectonic processes are suspect.

The sections describing the contrasts in chemistry of erupted rocks consequent upon changes in tectonic environment in areas of Africa–Arabia collision (Innocenti *et al.*), and collision and rifting in the Antarctic Peninsula (Tarney *et al.*), are important for our understanding of the interplay of tectonic and magmatic

processes. Further in this vein, the summaries of trace element and isotopic characteristics by Pearce and Hawkesworth provide a number of constraints on petrogenetic hypotheses attributing significant roles for the different tectonic elements involved in plate collision zones. Some of these constraints are not recognized by authors of other segments of the book.

Space devoted to the role of andesites in the evolution of continental crust and the

types of mineral deposits found in arc environments could have been expanded, but at least serves as an introduction to these subjects. Overall, then, this is not a book for the beginner, but a number of important contributions and ideas are included which make for fascinating reading. □

Richard J. Arculus is a Research Fellow in the Research School of Earth Sciences, Australian National University, Canberra.

Prospects for imprisoning solar energy

J.I.B. Wilson

Solar Cells: Operating Principles, Technology, and System Applications. By Martin A. Green. Pp.274. ISBN 0-13-822270-3. (Prentice-Hall: 1982.) £22.35, \$27.95. *Solar Cell Device Physics.* By Stephen J. Fonash. Pp.332. ISBN 0-12-261980-3. (Academic: 1982.) \$35, £23.20. *Photovoltaic and Photoelectrochemical Solar Energy Conversion.* NATO Advanced Study Institute Series, B69. Edited by F. Cardon, W.P. Gomes and W. Dekeyser. Pp.422. ISBN 0-306-40800-7. (Plenum: 1981.) \$49.50, £31.19. *Photochemical Conversion and Storage of Solar Energy.* Proceedings of the Third International Conference, Boulder, USA. Edited by John S. Connolly. Pp.444. ISBN 0-12-185880-4. (Academic: 1982.) \$34.50, £23.

THE photochemical and photovoltaic approaches to energy conversion are often linked, as a contrast to the mechanical-engineering-based photothermal converters of solar energy. Photochemical converters — including photosynthetic systems — and photovoltaic cells are similar in that both use high quantum energy from a radiant source to produce high-grade energy. Of the two technologies, however, only photovoltaic cells have reached technical maturity. The problems which still prevent the practical application of photochemistry are electrode degradation and too low a power conversion efficiency, but these drawbacks are countered by the ease with which liquid systems can be increased in area and by the possibility of combined energy storage, benefits which together offer the prospect of cheaper large-scale energy conversion than high-technology, solid-state solar cells.

An accurate impression of the state of play in both the solar chemical and electrical conversion techniques can be gained from the list of contents of these four books. Solar cells are now included in some undergraduate courses and so a number of teaching texts are appearing, of which this one by Green is the most comprehensive; the author uses few

equations and his book is one of the best introductions to solar cell technology for anyone with some knowledge of physical science. At the same time, priorities in solar cell research are changing from meeting the requirements of space technology for high-efficiency cells to an emphasis on lower costs; thus the basic physics is being examined in detail for possible alternative device structures and for additional solid-state phenomena which might add appreciably to power output. Fonash's research-level monograph considers *all* sources of photocurrent and photovoltage in a generalized model of solar cells, and it should provide the theoretical basis for any future development. The book can be recommended as a specialist theoretical text for semiconductor device engineers.

The other two books reviewed here are essentially the records of conferences and contain chapters based on lectures from experts on a range of specialized topics. This format usually appeals only to those who went to the meetings, or would have liked to attend, but both volumes do contain sections of broad appeal.

The NATO Advanced Study Institute, held in Ghent late in 1980, was restricted to energy systems which were not then commercially viable, or not even technically proven. Thus economics, the ultimate arbiter of success in this field, was not included. The Boulder conference had much more emphasis on biological aspects of photochemistry and on the attempts to produce an *in vitro* photosynthetic system, but there is some overlap with the contents of the NATO volume — even to the extent of containing an identical chapter by A.J. Nozik.

Since the first question to be asked about a new solar energy device is: "what is its efficiency?", the chapter by J.R. Bolton and colleagues in the Boulder conference proceedings should be prerequisite reading for those wanting to understand the limitations of the answers they receive. Although there is a strict thermodynamic limit to the solar energy conversion efficiency of a two-level quantum converter of about 31 per cent, various

researchers have paid particular attention to the ultimate physical limitations of selected converters, and Bolton *et al.* compare these calculations. Of especial value is the authors' comparison of similar quantities in chemistry, photovoltaics and thermodynamics. They also show that the two-photon process of photosynthesis provides a greater excess potential to overcome energy losses than the simple one-photon system. This may well be a significant indicator for attempts to build an efficient artificial system simulating natural photosynthesis.

Such attempts are described in some detail by M. Calvin in the critical review which opened the Boulder conference, and the fundamental problems are stated quite clearly: although effective optical absorption is achieved easily enough, conversion of the photon energy to a separable chemical product creates considerable difficulties. Usually two coupled systems are required for separate oxidation and reduction reactions in a closed cyclic system, and it has proved difficult to connect the two parts without excessive energy loss. This is well demonstrated by the breadth of attempts to photolyse water using sensitized absorbing aqueous liquids or photosensitive semiconductor electrodes. Also in the Boulder volume is a detailed description of *in vivo* chlorophyll research through which we may eventually assemble an artificial system; this section, however, will be difficult going for the non-specialist.

Of more general interest is the problem of physically containing chemicals and photoactive surfaces, since all photochemical systems, whether generating electricity or chemical fuel, rely on active interfaces to separate the various functions. In a green plant the sunlight-absorber, chlorophyll, is contained in a thylakoid membrane which apparently separates the initially reactive oxidation and reduction products. Some of the experiments in generating hydrogen and oxygen from water by photolysis use analogous artificial membranes to separate the two balanced reactions without impeding the transfer of electrons and protons. In the alternative "single cell systems", the redox catalysts generate hydrogen and oxygen together, but these must be separated afterwards.

This sort of research requires a blend of semiconductor physics and inorganic and organic chemistry; it relies heavily on the design of new chemical compounds with specific molecular structures. In this

context, the more successful sections in both of the symposium volumes come from the photochemical side. In contrast is the brief chapter by S. Wagner (Boulder conference) on photovoltaic cells which fails to put forward any interesting problems requiring a chemistry solution (for instance in the application of amorphous semiconductors).

Recent calculations suggest the possibility of a solar energy converter whose efficiency is limited only by that of an infinite array of separate absorbers with different bandgaps. This possible efficiency limit of 66 per cent (which also exceeds the 52 per cent efficiency of an ideal thermal converter) was only mentioned during discussions at the Boulder conference, but details have since been published by R.T. Ross and A.J. Nozik (*J. Appl. Phys.* 53, May 1982). The trick is to inject "hot" charge carriers from the semiconductor into the electrolyte before they can thermalize and come into equilibrium with the other free carriers, in a manner analogous to negative electron affinity photocathodes for vacuum photodiodes. In this way less of the excess photon energy above the semiconductor bandgap is lost as heat. In his chapters in the books under review, Nozik describes some systems which may be showing these hot-carrier injection effects, but a viable energy converting cell remains to be produced.

I have selected only a few topics for discussion here, but it should be clear that photovoltaic cells use sophisticated semiconductor technology to provide an efficient, reliable source of electrical energy: such research is now concentrated on reducing production costs without undue sacrifice of efficiency. Some of the cells which have this objective are covered in detail in the NATO book, but there is no description of conventional crystalline silicon cells and no account of amorphous silicon cells. Photochemical cells providing either chemical or electrical energy using solid-liquid interfaces are more difficult to make as well as to understand. Since they are mostly too inefficient or too short-lived because of electrode corrosion to compete with existing energy sources, their application is some years hence. In view of the overlap between the treatment of these devices in the NATO volume and in the proceedings of the Boulder conference, I prefer the latter for its wider spread of subjects, more attractive format and for its inclusion of photosynthetic processes. It is to be hoped that some of the current photochemical research which appears in this book as abstracts of contributed papers may attract sufficient further study to lead to the cheap, efficient solar energy converter which we all await. □

J.I.B. Wilson is Senior Research Fellow in the Physics Department, Heriot-Watt University, Edinburgh.

Science in ferment

S.J. Pirt

Prescott and Dunn's Industrial Microbiology, 4th Edn. Edited by Gerald Reed. ISBN 0-333-33630-5. (Macmillan, London: 1982.) £40.

FERMENTATION technologists need a source book of descriptions of the fermentation processes used in industry. For this purpose one of the first choices has for many years been Prescott and Dunn's manual, first published in 1940. Originally the book was a straightforward account of information of practical use in virtually every industrial fermentation, with emphasis on the traditional high-volume, low-cost products.

The development of our understanding of the scientific basis of fermentation has proceeded apace in the 20 years since the appearance of the third edition of Prescott and Dunn. This development probably constitutes the greater part of what is known today of the quantitative expression of fermentation processes, metabolic regulation and the application of genetic manipulation to fermentation. A further consequence of two decades of activity has been the unification of the diverse technologies depending on microbial action — brewing, dairy and other food fermentations, secondary metabolite production, vaccine production and biological purification of effluents. The provision of a common basis for all of these processes is one of the strengths of biotechnology.

All of this has meant that the updating of Prescott and Dunn presented a considerable challenge. The editor, appreciating the magnitude of this task, coupled with the increased range of fermentations, has solved the problem by restricting the scope of the fourth edition to fermentations related to the food and beverage industries. In addition, the original monograph style with two authors for the whole book has been replaced by the edited text with contributions by many authors. Essentially the only thing left of the original book is the title.

Although the aim has been to provide a descriptive account of the various fermentations, within this format there is wide coverage of the scientific basis on which the various technologies rest, and a good coupling of the science with the technology. There is a liberal sprinkling of important details which could be crucial to technologists who are concerned with the industrial operation of the processes; thus the authors seriously attempt "technology transfer". In all the tradition of Prescott and Dunn has been preserved, albeit with the inclusion of only some rather than all industrial fermentations. □

S.J. Pirt is Professor in and Head of the Department of Microbiology at Queen Elizabeth College, University of London.

The business of solar cells

Finance and Utilization of Solar Energy published by the Center for Innovation and Entrepreneurial Development at the University of California, Santa Cruz, contains the proceedings of a conference on the topic held at the Center in October of last year. The publication costs \$10 plus postage.

BOOKS RECEIVED

Chemistry

- BOCKRIS, J.O'M., CONWAY, B.E. and WHITE, R.E. (eds). *Modern Aspects of Electrochemistry*, No.14. Pp.661. ISBN 0-306-40845-7. (Plenum: 1982.) \$55.
- CHANEY, J.F., RAMDAS, V., RODRIGUEZ, C.R. and WU, M.H. (eds). *Thermophysical Properties Research Literature Retrieval Guide 1900-1980*. Vol. 2, Inorganic Compounds. ISBN 0-306-67222-7. (Plenum: 1982.) Np.
- COYLE, J.D., HILL, R.R. and ROBERTS, D.R. (eds). *Light, Chemical Change and Life: A Source Book in Photochemistry*. Pp.406. ISBN 0-335-16100-6. (Open University: 1982.) £6.95.
- DENNEY, R.C. *Key Definitions in Chemistry*. Pp.144. ISBN 0-584-10559-2. (Muller, London: 1982.) £3.95.
- FRANKEL, M. *Chemical Risk. A Workers' Guide to Chemical Hazards and Data Sheets*. Pp.96. ISBN 0-86104-362-6. (Pluto: 1982.) £1.95.
- FRIED, B. and SHERMA, J. *Thin-Layer Chromatography: Techniques and Applications*. Chromatographic Science Series, Vol. 17. Pp.320. ISBN 0-8247-1288-9. (Marcel Dekker: 1982.) SwFr.148.
- FURTER, W.F. (ed.). *A Century of Chemical Engineering*. Pp.463. ISBN 0-306-40895-3. (Plenum: 1982.) Np.
- GIANTURCO, F.A. (ed.). *Atomic and Molecular Collision Theory*. Pp.505. ISBN 0-306-40807-4. (Plenum: 1982.) \$59.50.
- KORYTA, J. *Ions, Electrodes and Membranes*. Pp.197. Hbk ISBN 0-471-10007-2; pbk ISBN 0-471-10008-0. (Wiley: 1982.) Hbk £17, \$39.60; pbk £7.90, \$18.75.
- KUMAKHOV, M.A. and KOMAROV, F.F. *Energy Loss and Ion Ranges in Solids*. Pp.296. ISBN 0-677-21220-8. (Gordon & Breach: 1981.) \$93.50.
- LESK, A.M. *Introduction to Physical Chemistry*. Pp.746. ISBN 0-13-492710-9. (Prentice-Hall: 1982.) Np.
- LÉVY, M., LE GUILLLOU, J.-C. and ZINN-JUSTIN, J. (eds). *Phase Transitions*. Cargèse 1980. Pp.462. ISBN 0-306-40825-2. (Plenum: 1982.) Np.
- MATJEVIC, E. (ed.). *Surface and Colloid Science*, Vol. 12. Pp.473. ISBN 0-306-40616-0. (Plenum: 1982.) \$55.
- MILLER, J.S. (ed.). *Extended Linear Chain Compounds*, Vol. 1. Pp.481. ISBN 0-306-40711-6. (Plenum: 1982.) \$52.50.
- PATAI, S. (ed.). *The Chemistry of Functional Groups Part I and Part II*. Part I ISBN 0-471-27871-8; Part II ISBN 0-471-27872-6. (Wiley: 1982.) £150, \$295 (Part I and II).
- RODRIGUEZ, F. *Principles of Polymer Systems*, 2nd Edn. Pp.575. ISBN 0-07-053382. (McGraw-Hill, Berkshire: 1982.) £22.50.
- SEYMOUR, R.B. and STAHL, G.A. (eds). *Macromolecular Solutions: Solvent-Property Relationships in Polymers*. Pp.233. ISBN 0-08-026337-2. (Pergamon: 1982.) £21.25, \$42.50.
- TAYLOR, J.B. and KENNEWELL, P.D. *Introductory Medicinal Chemistry*. Pp.202. UK Hbk ISBN 0-85312-207-5; pbk ISBN 0-85312-311-X; US ISBN 0-470-27252-X. (Ellis Horwood, Chichester/Wiley/Halsted Press: 1981.) Hbk £20; pbk £7.50, \$16.45.

Technology

- ATHERTON, D.R. *Nonlinear Control Engineering*. Pp.470. ISBN 0-442-30486-2. (Van Nostrand Reinhold: 1982.) £8.95.
- MILLHON, J.P. and WILLIS, E.H. (eds). *New Energy Conservation Technologies and Their Commercialization*. In 3 Volumes. (Not available separately). Proceedings of an International Conference held April 6-10 1981, Berlin. Pp.3269. ISBN 3-540-11124-7/0-387-11124-7. (Springer-Verlag: 1981.) DM 320, \$149.
- MORGAN, R.P. and ICERMAN, L.J. *et al. Renewable Resource Utilization for Development*. Pp.394. ISBN 0-08-026338-0. (Pergamon: 1981.) £19.75, \$39.50.
- NATIONAL RESEARCH COUNCIL. *Outlook for Science and Technology: The Next Five Years*. Pp.788. Hbk ISBN 0-7167-1345-4; pbk ISBN 0-7167-1346-2. (W.H. Freeman: 1982.) Np.
- SAHA, S.K. and BARROW, C.J. (eds). *River Basin Planning: Theory and Practice*. Pp.357. ISBN 0-471-09977-5. (Wiley: 1981.) £19.
- WILLIAMS, J.C. and BELOV, A.F. (eds). *Titanium and Titanium Alloys. Scientific and Technological Aspects*, Vol. 1. Pp.798. ISBN 0-306-40191-6. (Plenum: 1982.) \$195.
- WISEMAN, A. (ed.). *Topics in Enzyme and Fermentation Biotechnology*, Vol. 6. Pp.232. UK ISBN 0-85312-372-1; US ISBN 0-470-27304-6. (Ellis Horwood, Chichester/Wiley/Halsted Press: 1982.) £21.50.

Earth Sciences

- BONNEAU, M. and SOUCHIER, B. *Constituents and Properties of Siols*. Pp.496. ISBN 0-12-114550-6. (Academic: 1982.) £36.20, \$74.50.
- CUMBERLAND GEOLOGICAL SOCIETY. *The Lake District. Rocks and Fossils No. 4*. Pp.136. ISBN 0-04-554007-1. (Unwin, London: 1982.) £3.95.
- LEEDER, M.R. *Sedimentology. Process and Product*. Pp.344. Hbk ISBN 0-04-551053-9; pbk ISBN 0-04-551054-7. (George Allen & Unwin: 1982.) Np.
- MONMONIER, M.S. *Computer-Assisted Cartography. Principles and Prospects*. Pp.214. ISBN 0-13-165308-3. (Prentice-Hall: 1982.) £15.70.
- SALTZMAN, B. (ed.). *Advances in Geophysics*, Vol. 23. Pp.391. ISBN 0-12-018823-6. (Academic: 1981.) \$48.50.
- SIMPSON, I.M. *The Peak District. Rocks and Fossils*, No. 3. Pp.120. ISBN 0-04-554006-3. (Unwin, London: 1982.) £3.95.
- STIRLING, R. *The Weather of Britain*. Pp.270. ISBN 0-571-11695-7. (Faber & Faber: 1982.) £12.50.

VON RAD, U., HINZ, K. *et al.* (eds). *Geology of the Northwest African Continental Margin*. Pp.703. ISBN 3-540-11257-X/0-387-11257-X. (Springer-Verlag: 1982.) DM 110, \$51.20.

WILSON, G. and COSGROVE, J.W. *Introduction to Small-Scale Geological Structures*. Pp.160. Hbk ISBN 0-04-551051-2; pbk ISBN 0-04-551052-0. (George Allen & Unwin: 1982.) Hbk £10; pbk £4.95.

Applied Biological Sciences

- ALTSCHUL, A.M. and WILCKE, H.L. (eds). *New Protein Foods*. Vol.4, *Animal Protein Supplies. Part B: Food Science and Technology*. Pp.378. ISBN 0-12-054804-6. (Academic: 1981.) \$49.
- ASHTON, H. and STEPNEY, R. *Smoking: Psychology and Pharmacology*. Pp.222. ISBN 0-422-77700-5. (Tavistock, London: 1982.) £9.95.
- BECKER, F.F. (ed.). *Cancer 1. A Comprehensive Treatise*. 2nd Edn. *Etiology: Chemical and Physical Carcinogenesis*. Pp.714. ISBN 0-306-40701-9. (Plenum: 1982.) \$69.50.
- BENDER, A.E. and BENDER, D.A. *Nutrition for Medical Students*. Pp.380. ISBN 0-471-28041-0. (Wiley: 1982.) £16.
- BINDRA, J.S. and LEDNICER, D. (eds). *Chronicles of Drug Discovery*, Vol.1. Pp.283. ISBN 0-471-06516-1. (Wiley: 1982.) £24, \$43.
- BRUSHKE, A.V.G., VAN HERPEN, G. and VERMEULEN, F.E.E. (eds). *Coronary Artery Disease Today: Diagnosis, Surgery and Prognosis. Proceedings of an International Symposium held in Utrecht, May 1981*. Pp.455. ISBN 90-219-0503-5. (Excerpta Medica, Amsterdam: 1982.) \$107, Dfl.230.
- COHEN, J.S. (ed.). *Noninvasive Probes of Tissue Metabolism*. Pp.270. ISBN 0-471-08893-5. (Wiley: 1982.) £36.50, \$166.
- COMISION DE ONCOLOGIA. *Protocolos Terapeuticos Del Cancer De La Clinica Universitaria De Navarra*. Pp.354. ISBN 84-313-0724-2. (Ediciones Universidad de Navarra, S.A. Pamplona: 1981.) Np.
- CONSEILLER, C. *et al.* (eds). *Postgraduate Course 12th International Meeting of Anaesthesiology and Resuscitation 1980: Anaesthesia and Postoperative Care in Uncommon Diseases*. Pp.441. ISBN 9-0219-0473-X. (Excerpta Medica, Amsterdam: 1981.) \$74.50, Dfl.160.
- CONSEILLER, C. *et al.* (eds). *Postgraduate Course 12th International Meeting of Anaesthesiology and Resuscitation 1980: Volatile Halogenated Anaesthetics*. Pp.241. ISBN 9-0219-0474-8. (Excerpta Medica, Amsterdam: 1981.) \$55.75, Dfl.120.
- COSOFRET, V.V. *Membrane Electrodes in Drug-Substances Analysis*. Pp.362. ISBN 0-08-026264-3. (Pergamon: 1981.) £29.50, \$60.
- DE DATTA, S.K. *Principles and Practices of Rice Production*. Pp.618. Pbk ISBN 0-471-09760-8. (Wiley: 1981.) Pbk np.
- DENIS, L., SMITH, P.H. and PAVONE-MACALUSO, M. (eds). *Clinical Bladder Cancer. Proc. of an International Symposium presented at the Antwerp Medical Days held September 1980 in Antwerp, Belgium*. Pp.202. ISBN 0-306-40835-X. (Plenum: 1982.) \$29.50.
- DI CHIARA, G. and GESSA, G.L. (eds). *GABA and the Basal Ganglia. Advances in Biochemical Psychopharmacology*, Vol.30. Pp.252. ISBN 0-89004-752-9. (Raven: 1981.) \$28.

Psychology

- APTER, M.J. *The Experience of Motivation: The Theory of Psychological Reversals*. Pp.420. ISBN 0-12-058920-6. (Academic: 1982.) £16.80, \$34.50.
- BRENNAN, J.F. *History and Systems of Psychology*. Pp.374. ISBN 0-13-392209-X. (Prentice-Hall: 1982.) £15.70.
- GRAY, J.A. *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*. Pp.548. ISBN 0-19-852109-X. (Oxford University Press: 1982.) £27.50.
- WAGNER, D.A. and STEVENSON, H.W. (eds). *Cultural Perspectives on Child Development*. Pp.315. Hbk ISBN 0-7167-1298-X; pbk ISBN 0-7167-1290-3. (Freeman: 1982.) Hbk £14.80; pbk £7.60.

Sociology

- CAFAGNA, A.C., PETERSON, R.T. *et al.* (eds). *Child Nurtrance: Vol. 1. Philosophy, Children, and the Family*. Pp.377. ISBN 0-306-41003-6. (Plenum: 1982.) Np.
- GAUTIER, J.-Y. *Socioecologie: L'animal Social et Son Unvers*. Pp.267. ISBN 2-7089-8702. (Privat, France: 1982.) F.79.70.
- HERFURTH, M. and HOGEWEG-DE HAART, H. (eds). *Social Integration of Migrant Workers and Other Ethnic Minorities: A Documentation of Current Research*. Pp.265. ISBN 0-08-028957-6. (Pergamon: 1982.) £48, \$96.

History of Science

- ATHA, A. (ed.). *A Scottish Naturalist: The Sketches and Notes of Charles St John 1809-1856*. Pp.192. ISBN 0-233-97390-7. (André Deutsch: 1982.) £10.95.
- CANNON, J.T. and DOSTROVSKY, S. *The Evolution of Dynamics: Vibration Theory from 1687 to 1742. Studies in the History of Mathematics and Physical Sciences*, Vol. 6. Pp.184. ISBN 3-540-90626-6. (Springer-Verlag: 1981.) DM 98, \$45.70.
- CHRISTEN, Y. *Le Dossier Darwin*. Pp.252. No ISBN. (Editions Copernic, Paris: 1982.) F.75.